Workshop on Interdisciplinary Standards for Systematic Qualitative Research

Cultural Anthropology, Law and Social Science, Political Science, and Sociology Programs

National Science Foundation Supported Workshop

Report prepared by:

Michèle Lamont
Harvard University

Patricia White
National Science Foundation

Acknowledgements

We wish to thank James Granato, former NSF Political Science Program Director, Joane Nagel, former NSF Sociology Program Director, Stuart Plattner, former NSF Cultural Anthropology Program Director, and Christopher Zorn, former NSF Law and Social Science Program Director, for their help in planning and participation in the workshop, Karen Duke and Michelle Jenkins, NSF Social and Political Sciences Cluster staff members, for their administrative and technical support, Lauren Rivera and Sabrina Pendergrass**,** graduate assistants for their work with Professor Lamont on workshop and report preparations at Harvard University, Joane Nagel and Kristin Luker, Professor of Sociology and Jurisprudence and Social Policy, University of California, Berkeley for their thoughtful comments on the draft report, and the 23 workshop participants who took on the enormous task of representing their respective disciplines at the workshop through the submission and presentations of short papers prior to the workshop and comment and recommendations during and after the meeting. We also thank the 2007 graduate student cohort of the Department of Sociology at Harvard University and Camonia Long, Howard University graduate student, for their reactions to the report.

Workshop Participants & Attendees

Michèle Lamont, Harvard University, Workshop Organizer

Andrew Bennett, Georgetown University

Ted Bestor, Harvard University

Kathleen Blee, University of Pittsburgh

Don Brenneis, University of California, Santa Cruz

John Bowen, Washington University in St. Louis

John Comaroff, University of Chicago

David Collier, University of California, Berkeley

Colin Elman, Arizona State University

Wendy Espeland, Northwestern University

Linda Garro, University of California, Los Angeles

John Gerring, Boston University

James Granato, University of Houston

Wendy Griswold, Northwestern University

Jennifer Hochschild, Harvard University

Richard Lempert, National Science Foundation

Felice Levine, American Education Research Association

Jody Miller, University of Missouri, St. Louis

Joane Nagel, University of Kansas

Beth Rubin, University of North Carolina-Charlotte

Gery Ryan, Rand Corporation

Susan Silbey, Massachusetts Institute of Technology

Mario Small, University of Chicago

Kathleen Thelen, Northwestern University

Lisa Wedeen, University of Chicago

Susan Weller, University of Texas Medical Branch-Galveston

Patricia White, (Sociology) National Science Foundation

Alford Young, Jr., University of Michigan

Christopher Zorn, University of South Carolina

Executive Summary

On May 19-20, 2005, a workshop on *Interdisciplinary Standards for Systematic Qualitative Research* was held at the National Science Foundation (NSF) in Arlington, Virginia. The workshop was co-funded by a grant from four NSF Programs—Cultural Anthropology, Law and Social Science, Political Science, and Sociology—to Dr. Michèle Lamont, Harvard University. Professor Lamont was assisted in organizing the workshop by representatives from each discipline who coordinated group reports. The Cultural Anthropology group was co-chaired by Drs. Ted Bestor (Harvard) and Gery Ryan (RAND); Law and Social Science by Dr. John Bowen, (Washington University, in St. Louis); Political Science by Andrew Bennett (George Washington University); and Sociology by Dr. Kathleen Blee (University of Pittsburg).

It is well recognized that each of the four disciplines have different research design and evaluation cultures as well as considerable variability in the emphasis on interpretation and explanation, commitment to constructivist and positivist epistemologies, and the degree of perceived consensus about the value and prominence of qualitative research methods. Within this multidisciplinary and multimethods context, twenty-four scholars from the four disciplines were charged to (1) articulate the standards used in their particular field to ensure rigor across the range of qualitative methodological approaches;[1*] (2) identify common criteria shared across the four disciplines for designing and evaluating research proposals and fostering multidisciplinary collaborations; and (3) develop an agenda for strengthening the tools, training, data, research design, and infrastructure for research using qualitative approaches.

Prior to the workshop each participant prepared a short paper addressing the three topics (qualitative research standards, evaluation criteria, and agenda of future opportunities and needs). During the workshop, breakout sessions were held where each disciplinary group was asked to discuss and list the field's major qualitative research standards. These lists were then shared with the full workshop. Two categories of qualitative research standards were identified—(1) "shared" or key standards relevant to all four disciplines; and (2) standards judged essential by two or three disciplines. Participants also provided recommendations for improving qualitative research, training and infrastructures and identified promising areas of research in the four disciplines that would benefit from study using qualitative research approaches.

The workshop report is organized into four sections: Qualitative Research Design and Methods; Standards for Qualitative Research across Disciplines (in Anthropology, Law and Social Science, Political Science and Sociology); Recommendations for Producing Top Notch Qualitative Research; and Promising New Research Areas and Topics. Written commentary prepared by workshop participants is included in the Appendix. What follows is a brief summary of each.

1*Methodological approaches include ethnography, historical and comparative analysis, textual and discourse analysis, focus groups, archival and oral history, observational studies, interpretation of images and cultural materials, and unstructured and semi-structured interviews.

Qualitative Research Design and Methods

Workshop participants discussed the strengths of qualitative methods, standards qualitative research shares with quantitative research, and standards that are unique only to qualitative research. The major strength of qualitative research was judged to be the rich range of methodological tools available to study meaning, social processes, and group variations. The sample sizes and strategies of qualitative and quantitative approaches to research were contrasted. Qualitative research stresses in-depth contextualization, usually with small sample size. Qualitative research sampling techniques, while nonrandom, are usually attentive to demographic and theoretical dimensions. The combination of small and in-depth samples chosen for theoretical relevancy allows qualitative research findings a degree of significance or generalizability beyond individuals or single cases and provide opportunities to demonstrate rigor in all phases of a qualitative research project.

"Shared" Criteria for Designing and Evaluating Qualitative Research Across Disciplines

Workshop participants agreed that the four disciplines shared several standards for designing and evaluating high quality qualitative research. All value projects that:

- Situate the research in appropriate literature; that is, the study should build upon existing knowledge
- Clearly articulate the connection between theory and data
- Describe and explain case selection; why particular sites, participants, events, or cases are chosen
- Pay attention to alternative explanations and negative cases
- Operationalize constructs and describe expected findings
- Provide clear and detailed descriptions of both data collection and anticipated data analysis techniques: specify what counts as data, how the researcher will go about obtaining data and analyzing it
- Describe the intellectual, social, and political significance of the research
- Discuss generalizability or significance beyond the specific cases selected
- Specify the limitations of the research and anticipate potential reviewer objections
- Discuss the preparation of the researcher for the proposed project in terms of:
    - Cultural fluency
    - Language skill
    - Appropriate methodological/technical training
    - Knowledge of particular research context

Recommendations for Producing Top Notch Qualitative Research

Participants established that qualitative research could be enhanced by increased investments in education, training, and infrastructure. For example, NSF could enhance qualitative research through increased funding for students and scholars as well as publicizing its commitment to supporting high quality qualitative proposals. Support for pre-dissertation support, small pilot study grants for faculty, and student training opportunities through professional associations would likely be most effective.

Promising New Research Areas and Topics

Each disciplinary working group articulated new and exciting research areas that would benefit immensely from qualitative research or are taking advantage of qualitative research approaches, and in the process opening up new avenues of understanding. Social and cultural anthropological studies of responses to climate change, natural resources management, genetically modified food, food scarcity, and the global food trade would be greatly enhanced by being subject to high quality systematic qualitative analyses. Law and Social Science scholars are conducting interdisciplinary research on conservation, intellectual property and medicine by pairing with social studies of science and technology researchers using a range of qualitative approaches. In Political Science the rich contextualization of qualitative research could be used to more fully understand the rise of religious movements, relations between racial/ethnic and class identities and political actions, and policy-making processes, especially across institutions. Finally, most topics that sociologists study are amenable to both qualitative and quantitative analyses, but qualitative methods are particularly useful for studying timely topics such as group identities and boundaries; globalization at the micro level; race, class, gender, and age and health outcomes; and social and cultural meanings of food and obesity.

Table of Contents

Acknowledgements.....2

Executive Summary.....3

Background.....8

Qualitative Research Design and Methods: Strengths, and Shared and Unique Standards.....10
The Strengths of Qualitative Research.....10
Shared Standards of Qualitative and Quantitative Methods.....10
Standards Unique to Qualitative Research.....11

"Shared" Standards for Designing and Evaluating Qualitative Research.....13
Standards Shared by All Four Disciplines.....13
Standards Used by Two or Three Disciplines.....14

Recommendations for Producing Top Notch Qualitative Research.....15

Promising New Research Areas and Topics.....16
Anthropology.....16
Law and Social Science.....16
Political Science.....16
Sociology.....17

Appendices.....19

Appendix 1. List of Workshop Participants.....19

Appendix 2. Workshop Agenda.....20

Appendix 3. Cultural Anthropology.....23
Group Report: Defining Excellence in Qualitative Research.....24
Papers Presented by Workshop Participants.....27
Gery Ryan, Rand Corporation, *What Are Standards Of Rigor For Qualitative Research?*.....28
John Comaroff, University of Chicago – *Notes on Anthropological Method—Mainly in the Key of E*.....36
Linda Garro, University of California, Los Angeles, *Structured Interview Methods in Anthropological Research*.....41
Susan Weller, University of Texas Medical Branch-Galveston, *Standards of Rigor in Anthropology*.....47

Appendix 4. Law and Social Science.....51
Group Report: Defining Excellence in Qualitative Research in Law and Social Science.....53
Papers Presented by Workshop Participants.....55
Don Brenneis, University of California, Santa Cruz, *Warranted Trust*.....*56*
Susan Coutin, University of California, Irvine, *Qualitative Research in Law and Social Sciences*.....*60*
John Bowen, Washington University in St. Louis, *Zones of Methodological Convergence in Qualitative Social Science Research*.....*64*
Wendy Espeland, Northwestern University, *Thinking About Standards*.....*65*
Jody Miller, University of Missouri, St. Louis, *The Status of Qualitative Research in Criminology*.....*69*
Susan Silbey, Massachusetts Institute of Technology, *In Search of Social Science*.....*76*

Appendix 5. Political Science.....83
Group Report: Defining Qualitative Standards in Political Science.....84
Papers Presented by Workshop Participantsl.....89
Andrew Bennett, Georgetown University, *Overview of Organizational Efforts on Qualitative Methods in Political Science*.....*90*
David Collier, University of California, Berkeley, *A Statistical Rationale for Qualitative Research*.....*94*
Colin Elman, Arizona State University, *Explanatory Typologies in Qualitative Analysis*.....*97*
John Gerring, Boston University, *What Standards Are (or Might be) Shared?*.....*107*
Jennifer Hochschild, Harvard University, *Conducting Intensive Interviews and Elite Interviews*.....*124*
Kathleen Thelen, Northwestern University, *Defining and Communicating Standards of Rigor for Qualitative Research in Political Science*.....*128*
Lisa Wedeen, University of Chicago, *Qualitative Methods in Political Science*.....*134*

Appendix 6. Sociology.....141
Group Report: Defining Qualitative Research Standards in Sociology..... 142
Papers Presented by Workshop Participants.....147
Kathleen Blee, University of Pittsburgh, *Defining and Communicating Standards of Qualitative Research Rigor in Sociology*.....*148*
Wendy Griswold, Northwestern University, *Multidisciplinary Standards for Systematic Qualitative Research*.....*152*
Michèle Lamont, Harvard University, *Preliminary Thoughts on Qualitative Research in Different Disciplines*.....*155*
Joane Nagel, University of Kansas, *Conceptions from a Former NSF Program Officer*.....*161*
Mario Small, University of Chicago, *Lost in Translation: How Not to Make Qualitative Research More Scientific*.....*165*
Alford Young, Jr., University of Michigan, *Coming Out from Under the Ethnographic Interview*.....*172*

Background

On May 19-20, 2005 the Cultural Anthropology, Law and Social Sciences, Political Science, and Sociology Programs, and at the National Science Foundation (NSF) convened a workshop on *Interdisciplinary Standards for Systematic Qualitative Research*. The workshop builds upon and extends the foci and goals of a 2003 NSF Sociology Program report on the *Scientific Foundation of Qualitative Research*. In particular, the *Interdisciplinary Standards Workshop* was organized to encourage disciplinary representatives to articulate not only discipline specific standards, but to assess the appropriateness of standards from other disciplines for the evaluation and design of research projects. This was a major challenge since each discipline has a different research tradition and evaluative culture and considerable variation in emphases placed on interpretation and explanation, commitment to constructivist and positivist epistemologies, and degree of perceived consensus about the value and prominence of qualitative research methods. Within this multi-disciplinary and multi-qualitative method context, twenty-three scholars representing four disciplines – Cultural Anthropology, Law and Social Sciences, Political Science, and Sociology – were charged with three major tasks. The first was to articulate the standards used for rigor in qualitative methods across the range of methodological techniques, comprising ethnography, historical and comparative analysis, textual analysis, focus groups, archival and oral history, observational studies, interpretation of images and cultural materials, and unstructured and semi-structured interviews. Second, they were to identify common criteria that can be shared across the different disciplines for designing and evaluating research proposals and fostering multidisciplinary collaborations. And, third they were charged to develop a research agenda for strengthening the tools, training, data, research design, and infrastructure for research using qualitative approaches.

In order to facilitate discussion on a set of common questions and issues, each participant was asked to prepare a short paper prior to attending the workshop that responded to the following questions.

- What are the standards of rigor in your discipline (address methods with which you are most familiar)?
- How might these standards of rigor be communicated to or applied in other disciplines (i.e., common criteria for designing and evaluating research)? What are major areas of divergence between your own discipline and other social science disciplines? How might these be addressed?
- What are the most pressing issues of research design and methods facing qualitative research projects? Any suggestions for solutions?
- What areas or topics are most promising for investigations using qualitative methods?
- What areas of promising qualitative research are most likely to foster multidisciplinary projects?
- What is needed to strengthen tools, training, data, research design, and infrastructure for conducting qualitative research?
- What single advance in qualitative research methods or approach would contribute most to strengthening qualitative research?

These short commentaries served as the basis for workshop discussions. These discussions in turn were used to identify standards to use to design and evaluate qualitative research in the four disciplines.

Workshop participants also provided group reports assessing the state of qualitative research in each discipline (and in some cases sub-specialty) represented. Finally, participants generated a list of suggestions on how to effectuate improvements in training, education, and infrastructure for qualitative methods, and suggested substantive areas that might benefit tremendously from enhancements in qualitative methods. These recommendations reveal important differences in the four disciplines' epistemological and evaluative cultures. They also demonstrate that while standards might be discipline-specific, they also have a high degree of overlap with standards used by scholars with different disciplinary qualitative research and training experiences.

Qualitative Research Design and Methods: Strengths, and Shared and Unique Standards

The Strengths of Qualitative Research

Qualitative research encompasses a set of rich methodological tools which can be used either on their own or in tandem with quantitative data gathering techniques to explore a wide range of substantive problems. Such techniques, including interviews, archival research, and ethnography are particularly well-suited for examining complex social structures, processes, and interactions that require consideration of numerous dimensions and levels of analysis. In particular, qualitative methods can be used to explore micro-social phenomena as well as the cultural understandings actors bring to social experience, interactions, and institutions. Moreover, they are particularly valuable for unraveling the mechanisms underlying causal processes, especially those that occur over time.

Qualitative research fruitfully can be applied at all stages of the research cycle, not just for theory generation and hypothesis formulation. Although qualitative methods are rich methods for performing such tasks, they are also well suited for analytical aims conventionally considered to reside in the realm of quantitative modes of research, including refining or challenging existing theories and generating and testing hypotheses. They can be used to provide rich, "thick" description of particular cases and perform comparisons across research settings. They are also ideal for "process tracking" or for allowing the researcher to discern how processes emerge and evolve. Qualitative research enables scholars to gather detailed data about the experience of individuals within social contexts in a way that surveys conventionally cannot. Qualitative approaches allow for the inclusion of subjective experience and cultural sense making that play a vital role in understanding all facets of social life. Quantitative methods are excellent for testing whether or not a model accurately describes a social reality, but qualitative methods are superior for generating a model in cases where there is no clear starting point, for example, for studying how social groups define lines of distinctions or how social groups draw lines of distinction between themselves and others in different social contexts.

Other strengths of qualitative work are the flexibility and recursivity it affords researchers that are made possible by a closer engagement to people and groups being studied. A characteristic of much quantitative research is that it presumes to know in advance what the relevant "variables" are, and the boundaries around these variables. What looks rigorous in advance of research may in fact be less rigorous in its execution. Many forms of qualitative research permit the researcher to adapt to learning as they become more knowledgeable about the social context they are studying. If in the process of doing interviews, for example, one learns that things that were expected to be important to respondents were not, one can adjust and look for evidence to help understand why. This recursivity can be methodological (if one line of questioning does not work, try another), theoretical (if predicted patterns do not emerge, or conditions unelaborated in theory seem important, then think about how to develop the theory or look elsewhere), or practical (if you do not get access to one group, you try another).

Shared Standards of Qualitative and Quantitative Methods

Qualitative and quantitative methods share a number of important standards for evaluating research. Demonstrating rigor in all phases of a research project, from study design to data analysis, is essential to

producing excellent research in each methodological tradition. When using either method, the researcher should:

- Articulate a clear research question that is framed in terms of the relevant theoretical and methodological literature both within and beyond their particular discipline
- Define and operationalize key constructs and specify expected relationships between concepts
- Choose the type and source of data that will enable the researcher to answer the research question
- Demonstrate the intellectual, social, or political significance of the project – why the project is important and will have an impact beyond a limited circle of scholars
- Undertake systematic and thorough data collection, gathering an abundance of evidence and triangulating with multiple data sources and data types when possible
- Provide a careful articulation of the connection between theory and data
- Conduct systematic and thorough analysis of data, specifying the particular strategies used to identify patterns and relationships in the data
- Pay close attention to negative cases and explore alternative explanations when available

Because the relationships explored in qualitative research tend to be more complex and less straightforward than those examined by traditional survey research, qualitative researchers have to be sensitive to the need to convince reviewers that the work they are proposing will be rigorous. First and foremost, the scholar should provide a clear description and explanation of case selection – why particular sites, participants, events, or cases were chosen. Such selection should be driven by theoretical considerations rather than by convenience. Moreover, the scholar should possess the relevant skills and knowledge necessary to complete the study. These include cultural fluency, or knowledge about the specifics of the field site or chosen population(s), language skills, and appropriate methodological training (i.e., coursework in qualitative data analysis).

Standards Unique to Qualitative Research

Small-Samples Can Sometimes Yield Big Insights

Despite these shared standards of evaluation, there are some important points of departure that scholars should take into consideration when assessing qualitative research. In particular, case selection and sampling procedures tend to be employed differently in qualitative studies. Sample size is a function of the number of theoretical constructs involved in a project and the amount of agreement in responses. "Small-N" research design is oftentimes appropriate in qualitative research. If the cases are appropriately chosen with regard to theoretical factors and compared, they can yield unique insights by revealing regularities between categories of cases that may escape large sample studies. By design, large scale studies must keep surveys relatively short and although they may provide representative information, they do so by sacrificing depth or detail in questions and responses. Moreover, by thoroughly examining a small number of cases, the researcher may explore in-depth the contextual dimensions that influence a social phenomenon. Attention to such environmental and situational factors is often downplayed in larger-scale studies that often favor linear analysis and flatten out variegated social patterns even though they may be characteristic of social processes.

Systematic Sampling Can Still be Scientific, even if it is Not Random

Sampling procedures need to be evaluated according to the purpose of the project. As highlighted by Mario Small's paper (page 175-182), while random sampling is often not necessary for qualitative studies, systematic sampling is (see Jeffrey Johnson, Selecting Ethnographic Informants" *Sage Qualitative Methods Series 22*, 1990). As Small notes, random sampling does not necessarily decrease respondent bias and may even result in less accurate data than more purposive methods of participant selection. Since the purpose of a qualitative study is to acquire new, more detailed knowledge on a topic, selection methods and interviewing styles need to be suited to that purpose. Snowball sampling allows the researcher to enter into networks of individuals and identify respondants that they might not otherwise be able to identify. However, participants tend to be more honest and willing to divulge personal information to researchers who have been validated by someone they know, enabling the researcher both to gather more accurate data and speak to individuals who otherwise may have declined to participate in research with a complete stranger. Furthermore, particularly in the case of expert and elite interviews, referrals can help the researcher pinpoint those participants who are most appropriate for the study at hand. Similarly, when dealing with hard-to-reach populations (i.e., sensitive or covert populations such as illegal substance users) referrals may also be the only means of actually identifying research participants. When performing referrals or any other type of targeted sampling, it is essential to ensure that the sample contains enough variation along key demographic and theoretical dimensions to draw conclusions beyond the particular individuals studied.

Generalizability to Population and Broader contexts and processes

In addition to differential standards of case selection and sampling procedures, the criterion of generalizability tends to take on more complex meaning in qualitative studies. Although all good qualitative research should inform us about how social processes occur beyond the particulars of the study data, the scale of generalizability can vary significantly from project to project and depends on the specific aims of the study at hand. Although larger scale qualitative projects may endeavor to generalize to large populations such as nation states or entire ethnic groups, many more seek to inform us about smaller groups or patterns of interaction that can have great significance for our understanding of social processes. Good proposals articulate clearly that the data gathered are meaningful beyond the particular cases, individuals, or sites studied and specify precisely why they are significant, to whom, and to which institutions and processes the findings can be generalized. These can range from subpopulations (i.e., particular minority groups) to types of interactions (i.e., criminal proceedings) to sets of institutions (i.e., admissions committees of elite universities).

"Shared" Standards for Designing and Evaluating Qualitative Research

Workshop participants identified criteria unique to their particular field. In the discussion of these field specific standards, it became clear that there were standards that were or could be shared across the four fields for the purpose of proposal development for evaluation at the National Science Foundation. This section of the report presents those "standards" shared by Anthropology (ANTH), Law and Social Science (LSS), Political Science (PS), and Sociology (SOC) and those explicitly shared by two or three of the disciplines. The standards fell in the four general categories of (1) project framing, (2) research design, (3) data analysis, and (4) writing and presentation. In some cases, standards may be shared across all disciplines, but were not mentioned in the context of the discussion. All standards do not necessarily apply to the full range of qualitative methods, and may be more relevant to a particular data gathering technique.

Standards Shared by All Four Disciplines

Framing of project. Devoting attention to how a research project is framed relative to existing scholarship on the topic is one of the most important determinants of a meritorious research project. A strong research frame:

- Articulates a clear research question
- Seeks to be well-situated in the relevant theoretical and methodological literature but also makes a novel contribution
- Provides a clear and compelling justification for why the research to be undertaken is exciting, important, and significant intellectually, socially, and/or politically

Research design. Well-designed research incorporates a:

- Strong interplay between theory and data – theory informs the project at all levels of inception, from the selection of research question and choice of research design to data analysis
- Detailed and theoretically informed justification for case selection and sampling procedure
- Thorough and transparent description of the full range of methods to be employed

Data analysis. A major tenet of data collection and analysis is to include information to demonstrate how researchers will consider negative or disconfirming evidence. Some disciplines required that alternative explanations be explicitly addressed.

Writing/Presentation. Research proposals should be written in a manner that is clear and logically coherent and can be understood by scholars across disciplines. This is as a major way to facilitate cross-discipline review and research.

Standards Used by Two or Three Disciplines

Sociology and Anthropology have considerable overlap in their standards for research framing and design and data analysis, and both valued the expertise and skills of the researcher. A meritorious research project must meet a number of basic standards.

Framing of project.

- Use appropriate level of theory
- Provide theoretical expectations (or preliminary hypotheses) based on knowledge of literature and cases, including discussion of expected relationship between key constructs

Research design.

- Have a strong interplay between theory and data. The theory must inform the project at all levels of inception, from the selection of research question and choice of research design to data analysis
- Articulate a detailed and theoretically informed justification for case selection and sampling procedure
- Demonstrate that the applicant has thought out how to handle contingencies in the field and is able to respond productively to unexpected changes in the context of research (This standard was also supported by the Law and Social Science participants.)

Data analysis. Thoroughly describe plans for data analysis, that is, how themes will actually be found, patterns identified, and comparisons made in the data. The simple mention of the use of qualitative analysis software packages is insufficient. The design and analysis should link clearly to answering the research question.

Other.

- Demonstrate the necessary knowledge and skills to complete the project, in particular cultural fluency and language skills
- Attend to the issue of replicability. Describe the methodology in sufficient detail and transparency to allow other researchers to check the results and reanalyze the evidence. (This standard was also shared by Political Science.)

Recommendations for Producing Top Notch Qualitative Research

Enhanced funding for research and training opportunities for qualitative research could facilitate academic excellence across disciplines and foster the development of exciting and insightful qualitative research projects which can have important impacts both within and beyond the academy. There also is a tremendous need for formal training in qualitative methods and research, and the development of qualitative research infrastructure to enable the sharing of data and methodological tools.

Training and Outreach. Organizations such as NSF that fund social science research can have a huge impact on enhancing the quality of qualitative research produced across disciplines through more heavily investing in training for scholars performing qualitative research as well as publicizing a commitment to supporting high quality qualitative proposals. The development of teaching materials for qualitative methods, concerted efforts to share resources across-disciplines, and funding opportunities for research in a variety of areas of qualitative research (such as textual analysis, process tracing, ethnographic work across disciplinary divides) would greatly advance qualitative research and methods.

Pre-dissertation Research Support and Funding of Pilot Projects. Significant preliminary research is normally required in order to prepare a proposal that meets the standards outlined in this report. Pre-dissertation grants for graduate students are needed to enhance the quality of dissertation research and proposals by support to explore hypotheses, specify proposed mechanisms, and gain cultural fluency in their chosen research settings. Pilot grants for faculty members at all levels can achieve similar objectives for larger scale projects and are a necessity in the case of many multi-disciplinary studies.

Partnerships with Professional Associations. Discipline-based professional associations could be supported to create training programs for those utilizing and teaching qualitative methods. Such programs could include summer institutes focused on qualitative methodology (i.e., IQRM—Institute for Qualitative Research Methods, the qualitative equivalent of ICPSR—Interuniversity Consortium for Political and Social Research and the NSF-sponsored Summer Research Design Institute for Cultural Anthropology), workshops for teachers of qualitative methods, and programs aimed at developing teaching materials for qualitative research. Training programs such as these could serve to equip the next generation of scholars with the methodological tools and analytical skills needed to perform rigorous qualitative research.

Standardization and Dissemination to Create Qualitative Data Infrastructure . Support of efforts to create common standard of how qualitative data are to be archived and guidelines for public release of data and documentation would yield far-reaching impacts through enabling broader availability and usage of the qualitative data collections it supports. This would build cumulative qualitative data and methods and a larger multi-disciplinary user community.

Promising New Research Areas and Topics

Qualitative research employs a range of methodological tools to advance scientific knowledge in the social sciences. Workshop participants identified a number of research areas and topics that they considered to have the potential to yield significant fundamental insights, findings, and understandings if subject to concentrated investigations using a range of qualitative research approaches.

Anthropology

Social and cultural anthropologists agree that supporting ethnographic and qualitative research on regional, ethnic, and national identities as they react to trends of globalization, in terms of economic change, political mobilization, regional warfare, and religious revitalization, hold considerable promise for providing a better understanding of topics that have significant policy implications. Of particular relevance are studies of:

- religious revitalization and religious conflict
- the nature of secularism as social ideology
- secularism and gender equality/discrimination
- nationalism, the state, and cultural production
- political economy of global/local production, distribution, consumption

The aforementioned topics as well as other current issues generate basic research questions that would benefit from qualitative research perspectives and methods. These include human responses to climate change; resource management; and adaptations to global food systems, including but not limited to: food borne diseases (mad cow disease--Bovine spongiform encephalopathy and SARS--severe acute respiratory syndrome) and genetic modifications; food scarcity; structures of global food trade; and resource depletion, for example, overfishing and habitat destruction. Also, research on the impacts of commercialized popular culture in transnational contexts (including tourism as a form of mass consumption); the social construction of local, regional, global markets; and circulation and commodity chains could be enhanced by the rich contextualized findings that tend to emerge from well designed ethnographic and qualitative analyses.

Law and Social Science

Law and Social Science scholars currently conduct studies of legal consciousness that are necessarily qualitative in nature, and legal scholars are now experimenting with new approaches to bring knowledge of "culture" together with knowledge of "law." These new comparative and interdisciplinary approaches provide greater depth and have the potential to significantly advance our understanding of legal cultures and jumpstart studies of law and emerging technologies. Three examples follow.

- While older studies looked for cultural principles underlying law, some more recent studies analyze how judges, lawyers, litigants and others reason and justify different normative systems. This approach can be critical when reasoning and justification diverge.

- Another exciting emerging area of legal research is the following of claimants over time in Iranian divorce courts to see how one bargains and interacts with judges "in the shadow of the law" invoking norms of fairness that do not appear in formal adjudications (and are not "Iranian cultural principle").

- New "cutting edge" domains include the study of new reproductive technologies and how actors bring norms from religion, law and morality to decisions concerning fertility enhancement, abortion and other reproductive issues. Law and social science scholars often pair with scholars of the social studies of science and technology to bring new interdisciplinary perspectives as was recently done in the case of studies of conservation, intellectual property and medicine.

Political Science

In general, there are few substantive research programs within political science to which qualitative methods cannot contribute. These methods are likely to be especially useful, however, in the development and testing of theories about phenomena that are relatively new, infrequent, or complex, or that combine some or all of these three characteristics.

- In American politics, qualitative research might fruitfully focus on the rise of religious movements, relations between racial/ethnic and class identities and political action, and policy-making processes, especially those that cut across different institutions**.** Qualitative approaches also could be used to provide more informative studies of what public opinion polls measure and how Americans engage in the complex process of political decision-making.
- In comparative politics, qualitative research on democracy and democratization could be expanded to move beyond minimalist definitions to include the ways in which everyday democratic practices operate within authoritarian regimes. Studies of democracy might also explore the various meanings of the concept in different cultural and historical contexts.
- In international relations, fruitful areas for qualitative research include the role of non-state actors, terrorism and other non-traditional forms of combat, financial crises, proliferation of weapons of mass destruction, the effects of global trade on local communities, and competition among and with emerging great powers.

Sociology

In sociology, most topics are amenable to both qualitative and quantitative research. However, there are some types of research questions and analytical objectives that are particularly well-served by qualitative methodology. These include: complex social structures, processes, and interactions; studies of the mechanisms underlying causal processes, especially over time; naturally occurring processes and phenomena of social life; studies that focus on questions of 'how' and 'why'; and the use of in-depth interviews to clarify findings from survey research. Particular areas that could benefit from greater qualitative research are studies of:

- scientific research and evaluation
- manifestations of globalization at the micro level, including as units of analysis cities, cultural practices, families, interpersonal relations, urban labor markets, and gender relations

- impacts of race, class, gender, and age on health and other social outcomes
- mechanisms that underlie patterns of inequality, group identity, and social inclusion/ exclusion, as they apply to race, gender, class, ethnicity, etc...
- consequences of war and social conflict on communities and identities
- conceptions of equality and inequality
- religious beliefs and political participation in America and elsewhere
- social and cultural meanings of food and the obesity epidemic

Appendix 1.

Workshop Participants & Attendees

Michèle Lamont, Harvard University, Workshop Organizer

Andrew Bennett, Georgetown University

Ted Bestor, Harvard University

Kathleen Blee, University of Pittsburgh

Don Brenneis, University of California, Santa Cruz

John Bowen, Washington University in St. Louis

John Comaroff, University of Chicago

David Collier, University of California, Berkeley

Colin Elman, Arizona State University

Wendy Espeland, Northwestern University

Linda Garro, University of California, Los Angeles

John Gerring, Boston University

James Granato, University of Houston

Wendy Griswold, Northwestern University

Jennifer Hochschild, Harvard University

Richard Lempert, National Science Foundation

Felice Levine, American Education Research Association

Jody Miller, University of Missouri, St. Louis

Joane Nagel, University of Kansas

Beth Rubin, University of North Carolina, Charlotte

Gery Ryan, Rand Corporation

Susan Silbey, Massachusetts Institute of Technology

Mario Small, University of Chicago

Kathleen Thelen, Northwestern University

Lisa Wedeen, University of Chicago

Susan Weller, University of Texas Medical Branch-Galveston

Patricia White, National Science Foundation

Alford Young, Jr., University of Michigan

Christopher Zorn, University of South Carolina

Appendix 2.

Workshop Agenda

Workshop on Interdisciplinary Standards for Systematic Qualitative Research
May 19-20, 2005
National Science Foundation

Sponsored by
Cultural Anthropology, Political Science, Sociology,
and Law and Social Sciences Programs

Thursday, May 19

8:30-9:00am Continental Breakfast

9:00-9:30 Opening Remarks
Wanda Ward, Deputy Assistant Director
Social, Behavioral & Economic Sciences Directorate
Peg Barratt, Director, Division of Social and Cognitive Sciences
Richard Lempert, Director, Division of Social & Economic Sciences
Michèle Lamont, Harvard University, Workshop Convener

29:30-10:15 Standards in Cultural Anthropology
Gery Ryan, Rand Corporation, Co-convener & session moderator
Ted Bestor, Harvard University, Co-convener
John Comaroff, University of Chicago
Linda Garro, University of California, Los Angeles
Susan Weller, University of Texas Medical Branch-Galveston

10:15-10:45 Brief comments and discussion

10:45-11:00 Break

11:00-11:45 Standards in Political Science
Andrew Bennett, Georgetown University, Co-convener & session moderator
David Collier, University of California, Berkeley
Colin Elman, Arizona State University
John Gerring, Boston University
Jennifer Hochschild, Harvard University
Kathleen Thelen, Northwestern University
Lisa Wedeen, University of Chicago

11:45-12:15 Brief comments and discussion

12:15-1:15 LUNCH

1:15-2:00 **Standards in Law & Social Science**

John Bowen, Washington University, Co-convener & session moderator
Don Brenneis, University of California, Santa Cruz
Susan Coutin, University of California, Irvine
Wendy Espeland, Northwestern University
Jody Miller, University of Missouri, St. Louis
Susan Silbey, Massachusetts Institute of Technology,

2:00-2:30 **Brief comments and discussion**

2:30-3:15 **Standards in Sociology**

Kathleen Blee, Co-convener & session moderator
Wendy Griswold, Northwestern University
Michèle Lamont, Harvard University
Joane Nagel, University of Kansas
Mario Small, University of Chicago
Alford Young, Jr., University of Michigan

3:15-3:45 **Brief Comments and Discussion**

3:45-4:00 **Break**

4:00-5:00 **Standards that Cross Disciplinary Boundaries—Common Criteria**

5:00 **Adjourn**

Friday, May 20

8:30-9:00am **Continental Breakfast**

9:00-10:45 **Promising Topics for Qualitative Research**

10:45-11:00 **Break**

11:00-12:30 **Resources Available and Needed for Strengthening Qualitative Research**

12:30-1:30 **Lunch**

1:30-2:15 **Critical Areas for Discoveries to Advance Qualitative Research**

2:15-3:30 **Developing a Research Agenda for Multi-Disciplinary Qualitative Research**

3:30 **Closing Remarks and Adjournment**

Appendix 3.

Cultural Anthropology

Defining Excellence in Qualitative Research: Groups Report for Anthropology

The group felt it useful to expand the existing guidelines on qualitative methodologies for applicants to the NSF Anthropology Program. We set out specific standards and suggestions in order to help applicants describe more clearly the key elements of their qualitative research design and methodology.

Framing of Project

The project proposal should state clearly which question(s) the research will answer. One of the strengths of qualitative methods is that they can be used to address a diverse set of aims. Consequently, "good" research questions for qualitative research can be exploratory, descriptive, or comparative; can seek to test models and/or hypotheses; or combine one or more of the above elements. Regardless of what type(s) of question the project seeks to answer, it is crucial to:

- **Demonstrate the intellectual, social, and/or political significance of the research.** The proposal should describe clearly how the study contributes to theory**,** substantive knowledge on the subject**,** methodological knowledge, and/or how it may have real world social or political impacts. Bear in mind that any research may have multiple arenas of significance. For example, a study of fishing communities in Mexico may simultaneously address theoretical issues in economic anthropology, important questions about Mexican social and political history, and problems of marine environments. In presenting the significance of the study, focus on the central interests of the audience/program to which you are submitting the proposal.
- **Specify, define, and operationalize terms.** Explain how you intend to use concepts like "class," "identity," or "subjectivity" and plans to identify, assess, and, if necessary, measure their relevance in the context of the research.
- **Explain clearly relationships among the key constructs**, **and the phenomena, to which the research is addressed.** If you mean to argue for relations of cause and/or effect, give the reasons for so doing. Note that it is not necessary to use the language of independent or dependent variables, but you DO need to state the objectives of the project, and the means by which you intend to realize them, in terms that are transparent to other scholars.
- **Situate the research question in existing literatures.** The proposal should demonstrate familiarity with range of relevant theoretical, methodological, and topical literatures, taking care to engage the full range of relevant methodological approaches (including, where appropriate, quantitative ones) that have been applied to this problem.
- **Use the appropriate level of theory**. Use theories of the appropriate level and range for the project. Take care not to write at too high a level of abstraction or generality. Note, in this respect, that it is NOT necessary, every time you undertake a research project, to re-write Durkheim, Weber, Marx, or Freud.
- **Scholarly generosity.** It is more productive to build on the accomplishments of other scholars than to offer negative critique purely for the sake of making a claim for the importance and novelty of your work. Also, where possible, avoid justifying the project primarily on putative lacunae in the literature on your sphere of interest. Do not personalize critical arguments and, to the greatest extent possible, evince a spirit of scholarly generosity.

.

Research Design: Data Collection

The research design and methodology portion of a proposal should describe case selection, sampling procedures, data collection plans, and data analysis schemes. In particular, the researcher should take extra care to:

- **Demonstrate an articulation between theory and data.** Theory should inform all levels of the research design and analysis. Recognize that ethnographic research is inherently a "multi-methods" approach, and that techniques must be stated clearly in relationship to specific questions/aims. Furthermore, ensure that data collection efforts are appropriate for research aims and objectives. (If you plan to make comparisons, you need data that are comparable!)
- **Justify case selection and sampling procedures.** Clearly describe the criteria for selection of sites, cases, informants, and/or events/processes. Why will you pick a particular site or informant? If your plans had to change because the situation in the field has changed, how would these selection criteria make it possible to adapt the project to changing circumstances? Moreover, describe the degree to which each of the above levels are typical or atypical – to what degree do sites, cases, informants, or events/process capture central tendencies, margins, and variation in the population? To what degree do they represent extreme or unique cases?
- **Describe full repertoire of methods proposed**. Do not presume that the reader is familiar with the methods you plan to use**.** Consequently, avoid jargon and cite specific references for techniques where appropriate (i.e., do not cite anthologies of research methods). Moreover, demonstrate familiarity with the strengths and limitation of chosen methods and draw comparisons with other potential methods. Specifically, why did you choose these particular methods and not others?
- **Strive for an abundance of evidence and triangulate when possible.** Try to balance experience-near vs. experience-distant data (e.g., distinguish between those things that an individual or community might experience as an aspect of daily life (for example, loss of landholdings because of tight credit) versus those phenomena that are more remote, abstract, or macro-level (national fiscal policies that affect agrarian credit).
- **Describe the 'flow' of the project – how one phase of research will connect to the next.** Describe the staging of the investigation as a clear sequence of steps (and logical progression) of different phases – multi-methods are often not simultaneous but build on one another in logical ways.
- **Describe in detail data analysis strategies.** Describe plans for finding themes, identifying patterns, and making comparisons. If conducting a descriptive analysis, explain how you recognize, describe, and account for range, central tendency, and variation, outliers. If building models, explain how you will identify key constructs, identify the relationships among constructs, and ensure that the models are representative of the data collected. If making comparisons, explain how you will identify the key dimensions on which the comparison will be based, how you will assess similarities and differences, and how the reader can be assured you have not biased the results. Furthermore, describe any post-data collection selection or data manipulation processes (e.g., data cleaning exercises, data management processes, and elimination of outliers).
- **Do not just cite software**. Explain what the software does and how it will advance research objectives; demonstrate that you know the software.

- **Address issues of falsifiability**. Specifically, what kind of evidence or negative cases would allow for a reader to falsify conclusions?
- **Discuss issues of representativeness and generalizability**. To what degree will this research tell about something larger, and what is that larger thing (e.g., a class, a region, or a demographic phenomenon?).
- **Anticipate reviewer objections and respond to them.** Describe the limitations of the study before reviewers do. In particular, describe what you have done to minimize such limitations**.** Additionally, if you have made hard choices on controversial issues, describe the controversy and explain why you made the choice

Necessary Resources and Research Schedule

Demonstrate that you possess the relevant skills, knowledge, and experience to complete the project and have carefully thought out how you will do so in a timely manner.

- **Cultural fluency.** Describe knowledge of the cultural and social setting(s) in which you will conduct the research, including language abilities, sensitivities to local concerns, and local framings of appropriate behavioral modes. In addition to preliminary site visits, "e-siting" or "e-contextualization" (systematic and critical examination of relevant on-line sources of information including, as appropriate, on-line communities and discussion groups in order to understand local concerns, and points of view) can be a very useful resource for keeping up to date with the chosen site.
- **Timeliness.** Provide a clear timeline that demonstrates how the research design can be completed in a timely fashion**.** The timeline should discuss how different phases of research—including different locations or field sites and different types of data collection—will be sequenced and integrated.

Cultural Anthropology
Papers Presented by Participants

What Are Standards Of Rigor For Qualitative Research?

Gery W. Ryan
RAND Corporation

First, I think it would be helpful to make explicit some of my basic underlying assumptions and definitions regarding: (1) What is qualitative research? (2) What are the goals and objectives of qualitative researchers? and (3) What is the range and scope of qualitative data collection and analysis methods? Answering these questions first makes it easier to address issues of standards of rigor and how they might be best applied to the broad range of social science research methods and investigations.

What is Qualitative Research?

The terms "qualitative data analysis" and "qualitative research" are mischievously ambiguous. Does qualitative data analysis mean "the analysis of qualitative data" or "the qualitative analysis of data"? And what specific aspects of "research" does the word "qualitative" modify? The confusion can be eliminated by clearly distinguishing between data and analysis. Figure 1, page 33, lays out the possibilities.

Cell *A* is the *qualitative analysis of qualitative data*. Interpretive studies of texts, like transcriptions of interviews, are of this kind. Investigators focus on and name themes in texts. They tell the story, as they see it, of how the themes are related to one another and how characteristics of the speaker or speakers account for the existence of certain themes and the absence of others. Researchers may deconstruct a text, look for hidden subtexts, and try to let their audience know—using the power of good rhetoric—the deeper meaning or the multiple meanings in it.

Cell *D* refers to *numerical or statistical analysis of numerical data*. Lots of useful data about human behavior come to us as numbers. Closed-ended questions in surveys produce numerical data. So do national censuses. Organizations, from businesses to charities to zoos, produce numerical data, too—data about the socioeconomic characteristics of people who use their products or services, data about how often they have to replace managers and secretaries, and on and on.

Cell *B* is the *qualitative analysis of quantitative data*. This can involve the search for patterns using visualization methods, like multidimensional scaling, correspondence analysis, and clustering. Cell *B* is also about the search for, and the presentation of, *meaning* in the results of quantitative data processing. It's what quantitative analysts do after they get through doing the work in cell *D*. It includes everything from the finding of regularities in a scatter plot to the interpretation of meaning and substantive significance of statistical tests. Without the work in cell *B*, cell *D* studies are sterile and vacuous.

This leaves cell C, the quantitative analysis of qualitative data. This involves turning words, images, sounds, or objects into numbers. Scholars in communications, for example, tag a set of television ads from Mexico and the U.S. to test differences in how older people are portrayed in the two countries. Political scientists code the rhetoric of a presidential debate to look for patterns and predictors of policies. Archeologists code a set of artifacts to produce emergent categories or styles, or to test whether some intrusive artifacts can be traced to a source.

Given the dichotomies above, workshop participants may want to consider the standards for data collection as well as for standards of data analysis.

What Are the Goals and Objectives of Qualitative Researchers?

Whether conducting quantitative or qualitative research, social scientists typically address at least one of four fundamental research objectives. The general questions associated with each are shown in Table 1, page 33.

1. In exploratory mode, the goal is to discover themes and pattern and to build initial models of how complex systems work. Whether they are investigative journalists tracking a story, archaeologists looking for new sites, ethnographers studying cultural groups, or grounded theorists studying how the elderly experience chronic illness, researchers doing exploratory work follow leads and hunches. They take a step forward, and then they backtrack, trying to uncover what is there, to experience the phenomenon they are studying as fully as possible, and to identify what is common and what is unique.
2. There are four basic types of descriptions: *thematic*, *case*, *group*, and *cultural*. *Thematic descriptions* present the concepts and themes identified in a corpus of text. *Case descriptions* involve a single case, and include the listing of typical events as well as the listing of idiosyncrasies and exceptions. In *group descriptions*, researchers describe a set of cases (a set of individual people, a set of churches, a set of rituals), noting how individuals are both similar to and different from each other and how the differences are distributed. Group descriptions can be qualitative or quantitative, or both, and vary widely in the amount of precision involved. Some cases call for a broad, sweeping description of a phenomenon. In other cases, we want to know simply whether something is present or not, and if so, how much of it is there? In group descriptions, the amount of precision required influences sampling strategies. Finally, in *cultural descriptions*, researchers describe the culture in which the cases reside. Unlike group descriptions, which focus on the members of a group, cultural descriptions focus on the beliefs and practices *shared* by members of a group.
3. Comparisons can be made at the case and at the group level, and can be qualitative or quantitative or both. They can also be more precise or less precise. In some cases, a statement like "X is bigger than Y" is enough precision. At other times, we need to know exactly how much bigger something is and whether the difference between X and Y could be expected by chance.
4. Model testing includes the systematic assessment of a single hypothesis or a set of multiple, related hypotheses (i.e., a model). This can take place at both the case and the group levels. In fact, in many ways, the testing of models is a special type of comparison where cases and groups of cases are compared to an ideal set of relationships. For example, an investigator using grounded theory might use negative case analysis to assess how well a model fits a single case study. Another investigator, using classic content analysis might test a single hypothesis across many cases.

These basic research objectives are not inimical to one another. In fact, many projects involve exploration, description, comparison, and model testing. Some scholars rely on qualitative data for exploration and discovery, and then use quantitative data for testing models. Increasingly, though, research across the social sciences relies on a balanced, commonsensical mix of both kinds of data.

Given the degree to which research objectives and research methods are intricately intertwined, workshop participants may want to consider whether there should be standards for achieving specific types of research objectives or standards for conducting specific data collection and analysis techniques.

What Are the Techniques for Collecting Qualitative Data?

Over the last century, social scientists have invented hundreds of ways to collect qualitative and quantitative data. Figure 2, page 34, shows a rough taxonomy of data collection techniques. Data collection can be broken into three fundamental categories—techniques for indirect observation, techniques for direct observation, and techniques for elicitation—based on the degree to which we interact with the people being studied. The more interaction we have with people whom we study, the more we have to be concerned about reactivity, or response effects. A fourth category, mixed methods, includes combinations of the other three. For example, ethnography, participant observation, case studies and site visits are likely to involve indirect and direct observation, as well a variety of elicitation techniques.

Elicitation techniques can further be categorized into four fundamental types of interviews: unstructured, semi-structured, structured and mixed elicitation tasks (Figure 3, page 34). Unstructured interviewing can be either informal or ethnographic. Informal interviews resemble casual conversations and are characterized by a total lack of structure or control. Though ethnographic interviews often look and feel like casual conversations, both the researcher and the informant know they are conducting an interview. Unlike more structured interviewing, an ethnographic interview allows for longer questions and more in-depth probing.

Semi-structured and structured interviews are designed so that each informant is asked a set of similar questions. This is particularly important if the goal is to make comparisons across individuals or groups of individuals. The interviewer initiates the conversation, presents each topic by means of specific questions, and decides when the conversation on a topic has satisfied the research objectives. The respondent in the research interview is led to restrict his or her discussion to the specific questions posed by the interviewer.

Structured interviews are used to measure the magnitude of a phenomenon or to make more precise comparisons within and across groups. The power of such interviews is based on the assumption that if we systematically administer a standardized instrument to a group of people, then variations in their answers are more likely caused by differences among *them* rather by differences in the instrument to which they responded. Structured interviews include surveys and questionnaires, as well as domain elicitation techniques such as free lists, pile sorts, frame elicitation tasks, triad tests, and paired comparisons. Note that structured interviews may include open-ended and closed-ended questions that elicit both long and short qualitative and quantitative responses. Mixed strategies involve a combination of structured, semi-structured, and unstructured techniques. Each method has its advantages and disadvantages. For example, more structured techniques are better for making comparisons and less structured interviews may be more appropriate for early exploratory phases of research.

Workshop participants will need to decide whether to concentrate some of their effort on setting standards for how qualitative data are collected, or to focus on how such data are best analyzed once they are collected.

What Are the Techniques for Analyzing Qualitative Data?

Once the data have been collected, investigators want to: (1) identify themes and subthemes, (2) build and apply codebooks, (3) describe phenomenon, (4) make comparisons, and (5) build, display, test and validate models. The analytic options available to investigators for accomplishing these goals are staggering.

Figure 4, page 35, provides a general taxonomy of analysis techniques and research traditions. Like Tesch (1990), the taxonomy distinguishes between the linguistic tradition, which treats text as an object of analysis itself, and the sociological tradition, which treats text as a window into human experience. The linguistic tradition includes methods such as narrative analysis, conversation (or discourse) analysis, performance analysis, and formal linguistic analysis.

Within the sociological tradition, there are two kinds of written texts: 1) words or phrases generated by techniques for systematic domain elicitation; and 2) free-flowing texts, such as narratives, discourse, and responses to open-ended interview questions. Analysis techniques for words and phrases include componential analysis, taxonomies, and mental maps generated by such algorithms as multidimensional scaling, correspondence analysis and cluster analysis. The analysis of free-flowing text include: (1) word-based analyses such as key-words-in-context (KWIC), word counts, semantic network analysis; and (2) code-based techniques such as grounded theory, schema analysis, analytic induction, classic content analysis, content dictionaries, and ethnographic decision making to name a few. Like data collection techniques, each of these methods of analysis has advantages and disadvantages. Some are appropriate for exploring data, others for making comparisons, and others for building and testing models. Nothing does it all.

The breadth of analytical techniques makes establishing common standards of rigor quite challenging. For example, classic content analysts who seek to test hypotheses typically assess (and report) inter-coder reliability before moving forward with any additional analyses. In contrast, grounded theorist or schema analysts whose goal is to discover and describe social phenomenon rarely conduct systematic inter-coder reliability tests. To what degree should we expect different methods being used for different purposes to be held to the same standards?

What is meant by standards of rigor and what can we do to enhance them in the social sciences?

Below I offer a series of cautions and then a series of suggestions for how we might obtain standards of rigor across such a broad array of goals and methods.

Cautions

First, we need to avoid confusing research rigor with concepts such as measurement precision, quantification, and generalizability. These latter concepts are choices that must be made by each investigator in determining how to best meet his or her research objectives and are not something that should be inherently desired in-and-of-itself.

Second, we need to be cautious about making claims that some data collection or analysis techniques are "more" rigorous than others. If techniques are tools in a researchers' toolbox, then this is like saying that "A saw is better than a hammer because it is sharper."

Third, we need to be careful that we do not overly prescribe "standards" for specific methodological techniques. Methodological techniques are a class of researcher behaviors that share a common set of core properties but include a range of variations and nuances. The power of a research technique lies in its ability to be adapted to multiple research situations. Truncating the variability around a technique will only make the tool less useful.

Fourth, we need to avoid trying to link specific techniques to specific research goals. As tools, methods are a means to an end. It is surprising how such means can be adapted to serve many different goals. For example, I could easily imagine scenarios where paired comparisons could be used to explore, describe, compare, or test hypotheses.

Fifth, we need to stop associating standards and rigor only with confirmatory and hypothesis-driven research. I see no reason why we cannot set standards of rigor for exploratory and descriptive research as well. I suspect that some of the criteria will vary based on specific research objectives, while some of the criteria will cut across all types of research.

Suggestions

First, I would argue that rigorous research is research that applies the appropriate tools to meet the stated objectives of the investigation. For example, to determine if an exploratory investigation was rigorous, the investigator would need to answer a series of methodological questions such as: Do the data collection tools produce information that is appropriate for the level of precision required in the analysis? Do the tools maximize the chance of identifying the full range of phenomenon of interest? To what degree are the collection techniques likely to generate the appropriate level of detail needed for addressing the research question(s)? To what degree do the tools maximize the chance of producing data with discernable patterns? Once the data are collected, to what degree are the analytic techniques likely to ensure the discovery of the full range of relevant and salient themes and topics? To what degree do the analytic strategies maximize the potential for finding relationships among themes and topics? What checks are in place to ensure that the discovery of patterns and models is not superfluous? Finally, what standards of evidence are required to ensure readers that results are supported by the data? The challenge for workshop participants will be to identify what questions are most important for establishing research rigor and to provide examples of how such questions could be answered for those using qualitative data.

Second, I would argue that rigorous research must be both transparent and explicit. In other words, researchers need to be able to describe to their colleagues and their audiences what they did (or plan to do) in clear, simple language. Much of the confusion that surrounds qualitative data collection and analysis techniques comes from practitioners who shroud their behaviors in mystery and jargon. For example, clearly describing how themes are identified, how codebooks are built and applied, and how models were induced would help bring more rigor to qualitative research.

Third, we all need to become more familiar with the broad range of methodological techniques available to us. Social science has become methodologically parochial. Content analysts, grounded theorists, semantic network analysts, and analytic inductionists do not talk to each other. Cross-fertilization across methodological traditions, especially those that are dominated by a single discipline, is a rare event. Even more worrisome is the growing tendency for researchers to attack all problems with the same type of methodological hammer.

Fourth, reviewers of manuscripts and proposals need to be better selected, trained, and supervised. Reviewers should provide feedback as to whether the question or topic is of interest to the readership or funding agency and the degree to which objectives stated in the proposal or manuscript are met by the methods and data presented. Unfortunately there is a tendency for some reviewers to try to change the objectives of research to match their own methodological expertise. The classic example occurs when researchers conducting exploratory research are criticized for not using techniques that are more appropriate for hypotheses testing. The opposite, however, also occurs. On numerous occasions I have seen “qualitative” researchers insist that their colleagues use more unstructured data collection approaches even though these were less appropriate for their stated research objectives. Reviews would be more constructive if reviewers had had broader methodological experience and authors improved their ability to clearly express what they plan to do and why.

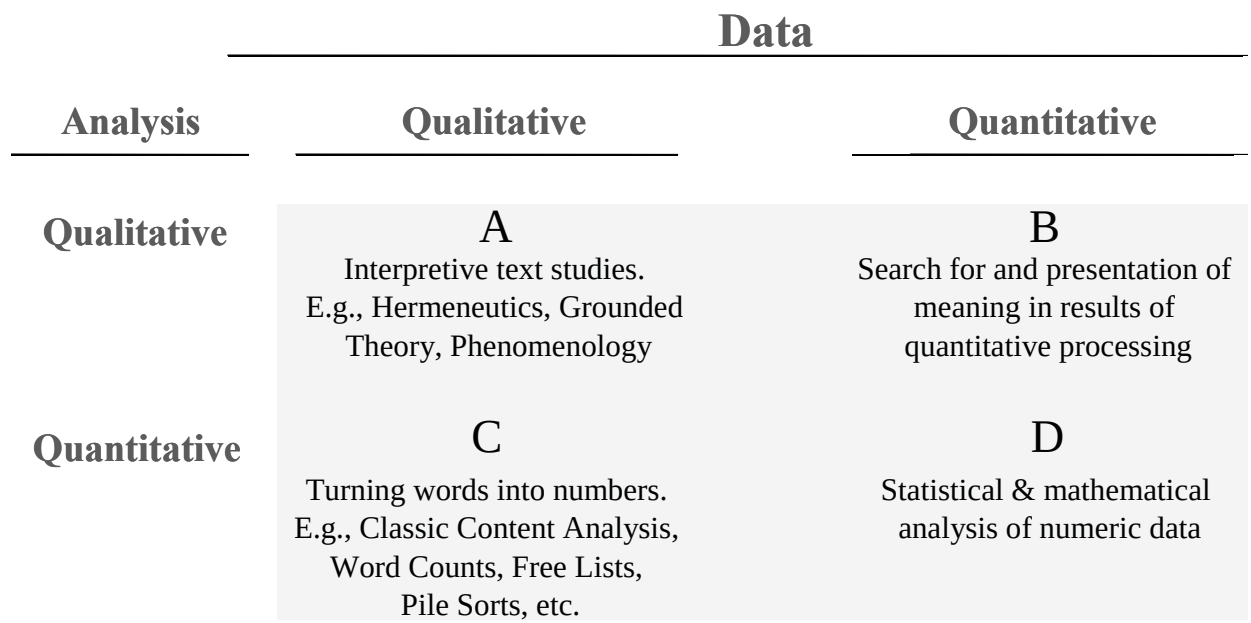
Fifth, there needs to be more methodological research on the basic techniques for identifying themes, building codebooks, marking texts, making comparisons and inducing models. There are many

different methods for accomplishing these tasks and it is unclear to what degree they produce different results. For example there is debate among grounded theorists about whether to follow the different techniques exposed by Glazer or Strauss. To my knowledge, there is no empirical evidence suggesting one is better than the other. Likewise, there are many different techniques of identifying themes in texts but again little to no empirical evidence about what works best and under what conditions.

Table 1. Goals of Qualitative Research

General Aim	Type	Questions
1. Exploration		What kinds of things are present here? How might these things be related to each other?
2. Description	Thematic	What does a theme look like?
	Case	What does a case look like?
	Group	What does a set of cases look like? Is a particular kind of thing (A) present or not? How much of that kind of thing (A) is there?
	Cultural	What does the culture look like?
3. Comparison	Case	How is Case X different from Case Y?
	Group	How is Group Xs different from Group of Ys?
4. Model Testing	Case	To what degree does a particular case conform to the proposed model?
	Group	To what degree does a group of cases conform to the proposed model?

Figure 1. Key Qualitative and Quantitative Distinctions



Adapted from: Bernard, H. R. (1996). "Qualitative Data, Quantitative Analysis." *Cultural Anthropology Methods Journal* 8(1): 9-11

Figure 2. Taxonomy of Qualitative Data Collection Techniques

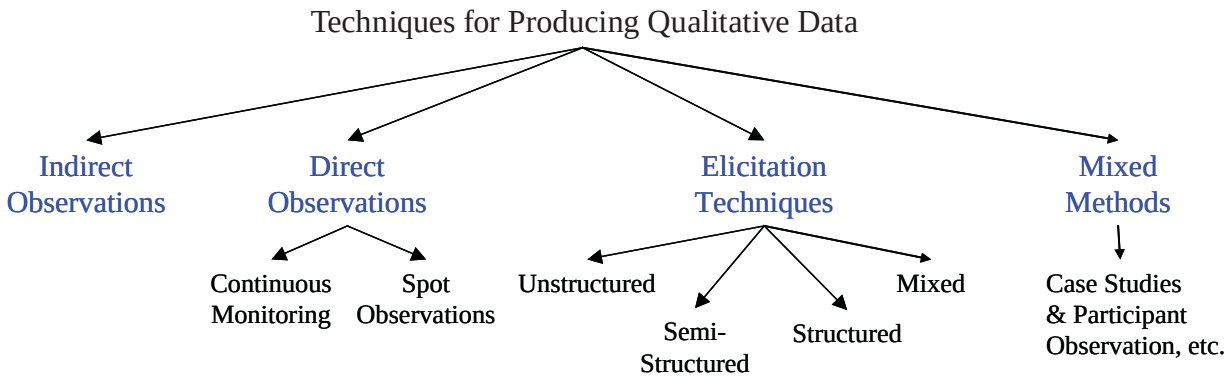


Figure 3. Taxonomy of Elicitation Techniques

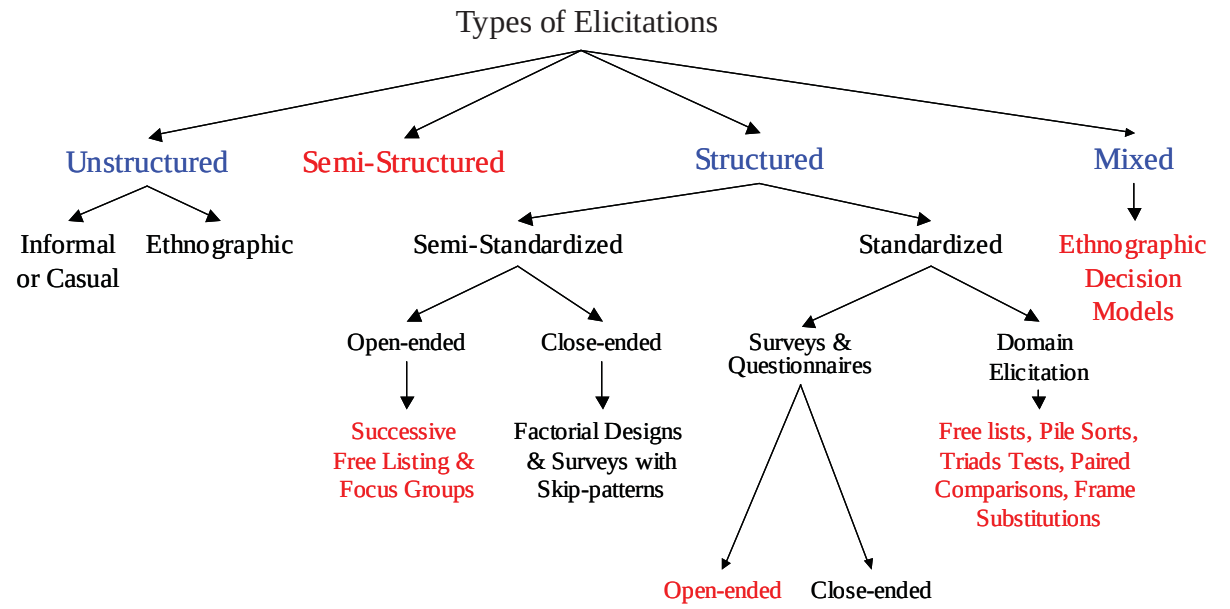
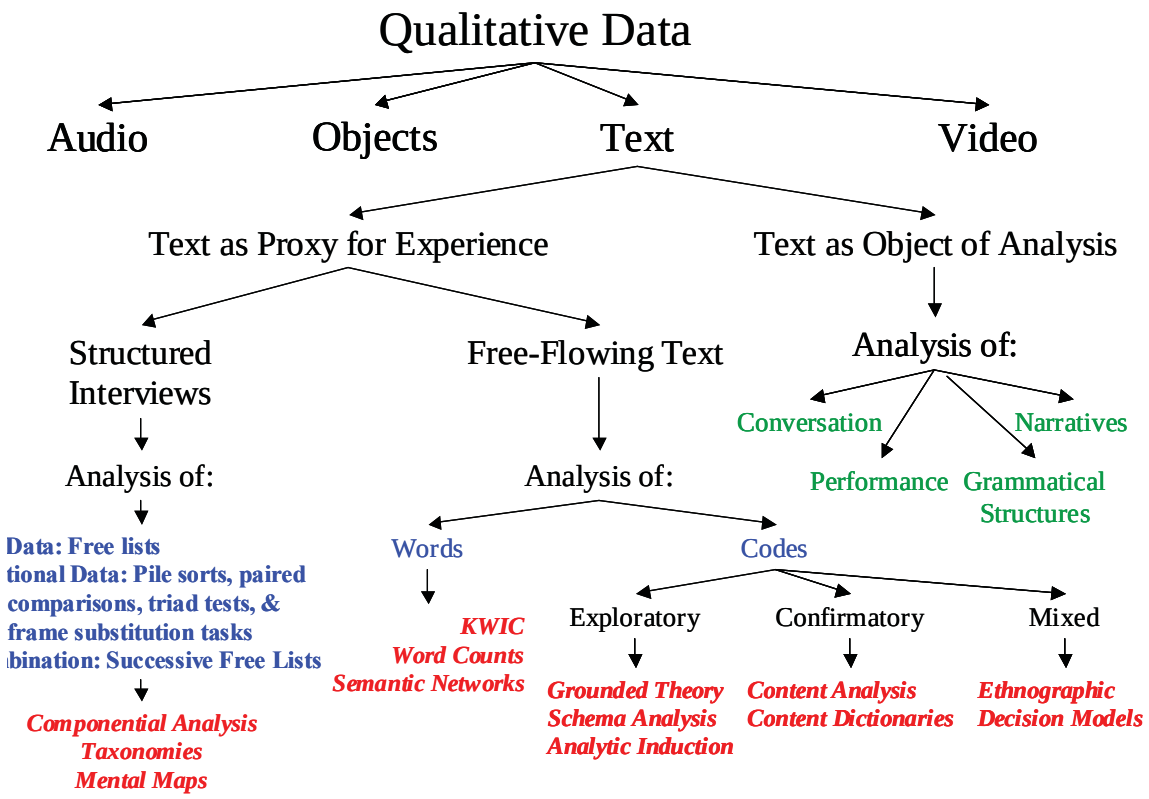


Figure 4. Taxonomy of Qualitative Analysis Techniques



Adapted from: Ryan, G. & Bernard, H. (2000). "Data Management and Analysis Methods." Pp. 769-809 in *Handbook of Qualitative Research, 2nd edition*, edited by N. Denzin and Y. Lincoln. Thousand Oaks, CA: Sage Publications.

Notes on Anthropological Method, Mainly in the Key of E

John Comaroff
Department of Anthropology
University of Chicago

"There is only one method in social anthropology, the comparative method – and that is impossible," – Sir Edward Evans-Pritchard[2]

Prolegomenon

It is exceptionally difficult to address questions of method for anthropology at large unless one reduces the field to a caricature of itself. To wit, the discipline – if anything so obviously polythetic, anything so unruly in its expansiveness, can be called a discipline at all – extends from the "hard" biological sciences and archaeological forensics through "comparative sociology" in the British tradition, technical linguistics of the various American schools, and ethno-everything in the formalist mode, to the "soft" hermeneutics of interpretive and historical ethnography. And this barely begins to exhaust the field. Even if we take Clifford Geertz's (1973) fashionable saw for the seventies, that anthropologists "do ethnography," it is difficult to pin down what this might mean in general, replicable terms; after all, Geertz's often imitated, widely-cited methodological approach – which was fleshed out only illustratively, never rigorously – was itself directed toward ideographic particularity and *away* from anything that might present itself as nomothetic, general, theoreticist. Since then, moreover, the discipline has become much more diverse, much less coherent; having departed the village, the reservation, the island, both its ends and its means are more contested than ever before.[3] Indeed, if, for David Lodge (1999:52), writing a novel is "like playing chess in three dimensions," doing anthropology nowadays is like playing it in four; the fourth being the terrain of the virtual, the electronic commons that ties even fairly remote social worlds into an expansive, if often exclusionary, global ecumene. In short, what follows is *a reductio,* if not *ad absurdam,* then certainly to the point of being a very provisional, very partial, very primitive sort of statement about contemporary ethnographic practice. It omits, for example, the question of historical method, critical though it clearly is to what we do much of the time; also the methodological challenges faced by anthropology as the scope and scale of its objects metamorphose.[4] It also refuses the idea that the qualitative social sciences can or should measure their methodological worth, or be evaluated, against the quantitative social sciences, whose techniques, their authority notwithstanding, are no less open to radical doubt (Comaroff and Comaroff 1992). The challenge for both lies in commensuration: how do we arrive *at once* at the necessary conceptual terms *and* at techniques of producing knowledge commensurate with the problems that we seek to address? This much is self-evident. It is here that the difficulties begin.

1 This much cited aphorism was never actually published; see Needham (1975:365).

2 Not only has its hallowed concept of "the field" been subjected to deconstruction (see e.g. Gupta and Ferguson 1997), but the dilemmas attendant on doing ethnography "in the modern world system" have been the object of considerable debate since the 1980s (see e.g. Marcus 1986; Bamford and Robbins 1997; fn. 3 below).

3 Jean Comaroff and I have addressed the question of method in historical anthropology elsewhere (1992); our take on methodical approaches appropriate to the ever more expansive anthropologies of the postcolonial, global age are discussed in another recent essay (2003).

Three or So Questions

On the matter of rigor, and its mortifications

Socio-cultural anthropology is often accused, to its mortification, of lacking all methodological rigor. Or worse yet, of having no discernable method at all – beyond its reliance on the self-assertively authoritative eye (or, more capaciously, the senses) of the ethnographer. Recall, in this respect, Levi-Strauss (1976:35): ethnography is a means of producing knowledge in which "[t]he observer apprehends himself as his own instrument of observation" (cf. also Foucault 1975). As Jean Comaroff (n.d.) has pointed out, its frank faith in the role of subjective experience in empirical investigation, and in the testimony of the lone investigator, has *always* been controversial: the anthropological mode of induction has repeatedly been accused of ineluctable ethnocentrism, of fetishizing difference, of celebrating imaginative idiosyncrasy, and of a cavalier disregard for replicability, refutability, or reliable accountability. To be sure, its most invoked technique, "participant observation," is often dismissed, at worst, as an oxymoron, at best, as a disingenuous synonym for "hanging out" with "the natives," whomever they may happen to be. Or, in slightly more dignified terms, engaging them in a "long conversation" (Bloch 1977)

Bracketing for now all the epistemic and ethical problems said to inhere in the methodological practices of anthropology – they are little different from those attendant upon the techniques of less reflexive disciplines and a lot less severe than those intrinsic to, say, modern journalism (see Steinberg 2002) – ethnography, done well, can and often does evince a great deal of rigor. And a high measure of refutability. For an ethnographic account to pass methodological muster, insisted Max Gluckman – founder of the Manchester School, which shaped post-WWII anthropological practice and developed some of its most innovative techniques – it ought to present primary data of sufficient quantity and depth to allow them to be reinterpreted from a theoretical perspective different from the one that produced them. This was the tacit standard to which much of British anthropology, famously empiricist in its orientations, held its practitioners until very recently.

I shall say more, in a moment, of the rigors and replicabilities of anthropological methods. First, though, a qualification: because their objects of research are diverse and protean, because the truths they pursue tend to be less nomothetic than those sought by most other species of social science, and because they are conditioned by the vernacular social realities and contingent human preoccupations with which they are confronted, it is always hard to lay out a set of ethnographic techniques in the abstract. Not, at least, without specifying the empirical problem to which they are to be addressed and the theoretical concerns that motivate, and are motivated by, that problem; in this sense, anthropology always rests on a dialectic between the deductive and the inductive, between the concept and the concrete, between its objectives and its subjects, whose intensions and inventions frequently set its agendas. The failure to grasp this may account in part for the autonomic dismissal of ethnography as unrigorous, unreplicable, unfalsifiable, and all the other (non-)u words with which it is regularly damned. More pragmatically, it means that its technologies have to be taken as a repertoire, an imaginative tool kit, whose various elements may be deployed, combined, and refashioned in an almost infinite variety of ways. It goes without saying that it is the ethnographer's obligation to explicate how and why s/he has deployed those elements in the way s/he has; for example, our effort to specify how we went about designing a theoretically-principled methodology with which to arrive at an account of, and to account for, the occult economies of postcolonial South Africa – economies made manifest in rampant witch-killing, zombie conjuring, AIDS related rape, financial frauds, and other complicatedly interconnected phenomena (Comaroff and Comaroff 2003).

The methodological toolkit available to the ethnographer – in addition, of course, to more conventional techniques, such as interviews, surveys, and focus groups of sundry sorts – begins with a number of "traditional" instruments whose uses are well-established: the extended case method and the social drama, for instance, both of which have well-elaborated, replicable procedures, and suggest ways of extrapolating general processes from singular events (see e.g. Epstein 1967; and, as applied in legal anthropology, Comaroff and Roberts 1981; Gulliver 1979); the recording of "natural discourse" (in court cases, public meetings, informal conversations, and ritual cycles), and of "native" exegeses on that discourse – all of which may be digitized for deposit and review – for purposes of both formal and content analyses; symbolic analyses of the kind pioneered by Victor Turner (e.g. 1967), which laid out a carefully ordered series of observational and interpretive procedures; the documentation at periodic intervals of "routine" activities, parsed by salient social categories (gender, generation, class, status, etc.); ethnomapping and the elicitation of other vernacular modes of representation, which, by virtue of being available for transcription and dissemination, may be subjected to scrutiny and re-analysis; the collection and aggregation of life-histories and the developmental cycles of collectivities, ranging from families to large-scale associations (see e.g. Goody 1958); network and transactional analysis (e.g. Mitchell 1969); and so on and on.

For anthropologists of my own generation, these remain some of the essential components of a solid spectrum of rigorous techniques that produce the kind of data out of which compelling accounts may be written: accounts that may be invoked in support or rejection of broad theoretical positions, of more specific explanations for social and cultural phenomena, of claims about the human predicament, past and present. Of late, they have been augmented by more literary and hermeneutic modes of descriptive-analysis, and by rather more impressionistic, aesthetic, reflexive approaches to the act of observation itself; also by attention to domains of human existence – from mass media and finance capital to the workings of the electronic commons and the dispersed practices of governmentality, to mention a few at random – hitherto not subjected to ethnographic methods. These can and have yielded exceptionally detailed, rigorous accounts of an extraordinary range of phenomena; but their rigor is not easily specified in programmatic terms. Indeed, they rely, for their persuasiveness, on their plausibility – itself often judged by virtue of their density, their imaginative scope, their capacity to bring a wide range of recognizable "facts" into a single descriptive-analytic purview – and their aesthetic composition.

Which leaves the question as posed: what are the standards of rigor in anthropology? In the abstract, there are none, none that are universally shared. Given that all method is mediated by theory and *vice versa*, our standards are, in the final analysis, determined contextually. If anything more general is to be said, it is that we tend to assess our techniques of knowledge production by the degree to which they yield data about which a cogent argument can be made in terms of prevailing conceptions of plausibility, persuasiveness, parsimony, density. Or, as Gluckman put it, the extent to which they yield accounts about which we may reasonably disagree, accounts that may reasonably be subjected to reinterpretation.

Communicating our Differences, Differentiating our Communications

The biggest challenge for anthropology is *not* to find more rigorous methods. As I have said, despite stereotypic caricatures to the contrary, ethnography practiced well – a qualification that applies to *all* methodology – is quite rigorous enough. Not only can it stand critical scrutiny as a mode of producing knowledge; it has long yielded truths of enormous insight and value, often to the discomfort of conventional Western wisdom. The challenge, in my own view, is to convince its practitioners that they owe it to themselves, and to their colleagues in other disciplines, to *explicate* their procedures

fully.[4] There is, if we are to be honest, a degree of high-handedness, even assertive contempt, among anthropologists, for speaking about our method: ethnography, to put it crudely, is what makes us different – or, rather, used to – and the rest of the social sciences can take it or leave it. Of course, "doing fieldwork, an extended spell of participant observation, remains a necessary rite of entry into our collegium (Ortner 1997:61); like all rites de passage, its mystique lies in *not* disclosing too much of its secret, even when the secret is that there is not much of a secret to it at all. Understandable though this might be as an exotic cultural practice, there is something self defeating to it. After all, on purely epistemic grounds, making an argument of commensuration for the relationship between an ethnographic account and the mode of producing it is an essential part of making a persuasive theoretical claim, or offering a persuasive explanation, for the phenomena under anthropological scrutiny. To be self-evident about all this: If an anthropologist wants, for example, to assay the view that a new religious movement has arisen in response to changes in local material conditions, her or his claim can only ever be as strong as is his or her demonstration that the new religious movement actually exists, that the alleged material transformations have actually occurred, and that the two things share a set of spatial, temporal, and experiential coordinates – all of which demands that we be shown, explicitly, the means by which these things have been established. Without this act of commensuration, no theory or explanation, however exquisite, imaginative, ingenious, will stand scrutiny. For all the fact that this is obvious, it seems rarely to be honored in the discipline. If we are to communicate about our practices, and convince other disciplines of their worth and their rigor, we have to begin with three principles: explication, explication, explication.

All the rest follows. The extent to which there are or will be common criteria for designing and evaluating research between anthropology and the other social sciences depends on how much the latter incorporate ethnography into their repertoires. And are prepared, reciprocally, to make the effort to understand what it actually is, what the e-word actually stands for. To the degree that they do – and many of them are, increasingly, if sometimes only at their margins – engaging in common substantive discussions about technique is more than appropriate, largely because anthropology *does* have a large and well-honed toolkit to offer.

Topics of Caprice and Corn: Promising Horizons for Qualitative Research

There are very few topics which *cannot* be illuminated by qualitative research, again well done. It depends how they are defined and formulated, what questions are asked of them, what it is about them that we wish to explain. Ethnography, these days, is being applied to an extraordinary range of things, from the caprice of futures markets to the science of genetically modified corn. And a great deal in between. Indeed, one way to answer this question as posed is to elicit an inventory of doctoral dissertation titles from any major anthropology department in the country – or, better yet, several of them. One thing becomes clear: in the rapid expansion of subject matter deemed suitable for ethnographic treatment, ethnography itself is undergoing an imaginative explosion, its horizons ever widened by its modes of knowledge production.

A closing thought. Perhaps the time has come to address the conundrum at the core of this workshop, well...methodically. And methodologically. How? By doing an ethnography of ethnography. If nothing else, it might prove that Evans-Pritchard was wrong all along. Anthropology has, from the first, had a plurality of techniques. What is more, comparatively speaking, its methodological tools have grown over the years. Indeed, far from being impossible, they open up all sorts of promising possibilities for the future.

4 I do *not* have in mind here the kind of reflexivity called for in the 1980s, which had it that the more an anthropologist revealed about her- or himself – as if such revelations are ever unmediated or unedited – the better placed the reader would be to assess her or his accounts of others. For the most part, this call became a prescription for a great deal of numbingly boring, self-serving prose and amateur auto-biography; it is a chapter in the recent history of the discipline better left closed for now.

References

Bamford, Sandra, and Joel Robbins, eds. (1997). *Fieldwork in the Era of Globalization*, special issue of *Anthropology and Humanism,* 22 (1).

Bloch, Maurice. (1977). "The Past and the Present in the Present." *Man (N.S.)* 13: 591-599.

Comaroff, Jean. (n.d.). "Ethnography: Colonizing Gaze or Humanizing Art?" In *New Dictionary of the History of Ideas*, edited by M. C. Horowitz. New York: Charles Scribner's Sons.

Comaroff, Jean, and John L. Comaroff. (2003). "Ethnography on an Awkward Scale: Postcolonial Anthropology and the Violence of Abstraction." *Ethnography,* 4:147-179.

Comaroff, John L., and Jean Comaroff. (1992). *Ethnography and the Historical Imagination.* Boulder: Westview Press.

Comaroff, John L., and Simon A. Roberts. (1981). *Rules and Processes: The Cultural Logic of Dispute in an African Context.* Chicago: University of Chicago Press.

Epstein, Arnold L. (1967). *The Craft of Social Anthropology.* London: Tavistock.

Foucault, Michel. (1975). *The Birth of the Clinic: An Archaeology of Medical Perception.* (Translated by A. Sheridan Smith). New York: Vintage Books.

Geertz, Clifford. (1973). *The Interpretation of Cultures: Selected Essays.* New York: Basic Books.

Goody, Jack, ed. (1958). *The Development Cycle in Domestic Groups.* Cambridge: Cambridge University Press.

Gulliver, Philip H. (1979). *Disputes and Negotiations: A Cross Cultural Perspective.* London: Academic Press.

Gupta, Akhil, and James Ferguson, eds. (1997). *Anthropological Locations: Boundaries and Grounds of a Field Science.* Berkeley: University of California Press.

Levi-Strauss, Claude. (1976). *Structural Anthropology, Volume 2.* (Translated by M.Layton). New York: Basic Books.

Lodge, David. (1999). *Home Truths. A Novella.* London: Penguin.

Marcus, George E. (1986). "Contemporary Problem of Ethnography in the Modern World System." In *Writing Culture: The Poetics and Politics of Ethnography,* edited by J. Clifford and G. Marcus. Berkeley: University of California Press.

Mitchell, J. Clyde, ed. (1969). *Social Networks in Urban Situations.* Manchester: Manchester University Press.

Needham, Rodney. (1975). Polythetic Classification: Convergence and Consequences. *Man (N.S.),* 10(3):349-69.

Ortner, Sherry B. (1997). "Fieldwork in the Postcommunity." *Anthropology and Humanism,* 22:61-80.

Steinberg, Jonny. (2002). *Midlands.* Johannesburg: Jonathan Ball.

Turner, Victor Witter. (1967). *The Forest of Symbols: Aspects of Ndembu Ritual.* Ithaca: Cornell University Press.

Structured Interview Methods in Anthropological Research

Linda C. Garro
Department of Anthropology
University of California, Los Angeles

Much of my research addresses the way concerns about illness enter into and are dealt with in everyday life. In this short piece, I discuss the use of qualitative structured interview methods within the context of research oriented around learning about the cultural domain of illness. I use the term "structured" or "systematic" to refer to interview approaches that, at least part of the time, ask the same or similar questions of all participants – here, I focus on interview approaches that ask the same question of all. As will be seen, the ability to compare responses across individuals facilitates an assessment of the extent of sharing and variability, both within and across cultural settings.

Drawing on my work, I situate the use of structured interview methods in relation to what Spradley (1980) has characterized as the "ethnographic research cycle." Agar (1996:62) provides a concise summary of this cycle, the concurrent process of data collection and analysis in ethnographic field research:

> In ethnography. . . you learn something ("collect some data"), then you try and make sense out of it ("analysis"), then you go back and see if the interpretation makes sense in light of new experience ("collect more data"), then you refine your interpretation ("more analysis"), and so on. The process is dialectic, not linear. Such a simple statement is so important in capturing a key aspect of doing ethnography.

In addition to facilitating "refinement" at the interpretative level, structured interview methods can serve to create a context where one is more likely to discover new information relevant to one's research goals. In my own work, the use of structured methods has also contributed to my ability to ask and address new questions that arose during the course of research.

I refer to research carried out at my two main fieldwork sites. The first site is Pichátaro, a town in the highlands of the west-central Mexican state of Michoacán where both Purépecha (Tarascan) and Spanish are commonly spoken (most of this work was carried out in conjunction with James C. Young). The second site is an Anishinaabe (Ojibway) community in Manitoba, Canada. At both sites, the research relied on structured data collection methods, more traditional ethnographic approaches, including participant observation, the collection of illness histories, and in-depth interviews. In both locations, fieldwork yielded many opportunities to learn from others through informal conversations about illness as well as to hear about and follow illness episodes in ways that enriched, corroborated, and challenged my ever-evolving understanding of how illness takes on cultural meaning and what people do when faced with illness. Such opportunistic learning from others is part and parcel of anthropological fieldwork.

Even at an early point in the research, simple, structured interview methods, like "free-listing," can create a context where the researcher is more likely to learn about topics of interest. In Pichátaro, individuals were asked to simply list the illnesses that they knew. This was then followed up with conversational inquiries about the characteristics of the illness, possible treatment actions, past occurrences of the illness in the household and course of treatment, and relationships with other illnesses. These conversations proved to be a rich source of information about illness and its treatment, information that guided and refined subsequent inquiries. Another free-listing task centered on treatment alternatives available to community members.

The interviews organized around the free-listing task provided the core for another structured interview – a term-frame substitution task. A term-frame questionnaire is useful for discriminating between and illustrating the nature of relationships among a set of named items. To construct the term-frame interview, a set of illness terms was selected from among those mentioned in the free listing task. Then a set of statements, again mainly taken from the free-listing interviews, were recast as question frames. Each illness term was placed in each frame and participants were asked whether the resulting statement was true or not. Although a qualitative interview format, the nominal responses are amenable to quantitative analysis. When a structured interview of this type is given orally, comments and reflections occasioned by the term-frame questions serve as another source of, at times quite illuminating, information for the researcher.

The description above, however, makes the task of constructing the structured interview itself sound more straightforward than it actually is. Although a term-frame interview is not intended to comprehensively cover cultural understandings about a given domain, it should be representative of the ways people in the community think about and talk about illness. Rather than being a "short-cut", considerable ethnographic grounding is required to construct a suitable term-frame interview, that is, one with the potential to afford insights about the cultural domain. And during the analysis phase, various techniques (e.g. hierarchical clustering analysis, multidimensional scaling analysis) can be used to discover patterning across responses (see Young and Garro, 1994, ch. 4; Garro 2000b). In my own work, I've tended to carry out multiple analyses of the same data, looking for areas of convergence and divergence. What is critical, however, is being able to link the overall patterning in the formal analysis with other ethnographic findings. In the Pichátaro study, there was considerable correspondence between the formal analysis and what people said and did in other contexts. Without going into details, in addition to providing converging evidence for our emerging interpretations, what the formal analysis contributed was a clearer understanding of the underlying conceptual distinctions that matter most to people in preventing and dealing with illness in everyday life. Further, at times, responses given in the structured interview revealed patterning that challenged our then current interpretations of local cultural understandings. For example, responses given in the structured interview provided an initial clue that local understandings of "contagion" bear no strong resemblance to biomedical germ theory but more typically draw on culturally relevant notions about the impact of environmental conditions on health. Thus, even though contagion is commonly used to refer to illnesses affecting many people at the same time, such illnesses are often explained as due to a large number of people being exposed to the same environmental conditions that lead to the illness (e.g., hot weather).

One question that arose during the course of fieldwork was whether local curers (*curanderas*) shared the same understandings about illness as non-specialists in the community (Garro 1986). The basic research question was whether the *curanderas* "validate their curing functions by commanding a specialized, esoteric body of knowledge about illness and its manifestations, or rather does their practice depend on specific skills in the implementation of a more or less shared body of knowledge" (Garro 1986: 352-353). A smaller, but representative, version of the term-frame questionnaire was created and given to a group of local women, half of whom were curers. Because the data permit direct comparisons, it was possible to test three hypotheses about the distribution of cultural knowledge about illness and its treatment: a) that there are no patterned differences in cultural understandings between curers and non-curers; b) that curers had an essentially separate system of knowledge distinctive from non-curers, and; c) curers form a systematic variant in that they agree more with other curers, but both curers and non-curers share the same conceptual organization of medical knowledge. The analysis supported the third hypothesis. Beyond the group level, it was possible to look at individual differences with regard to the extent to which each participant could be said to be in agreement with shared understandings about illness. As one might expect, those individuals in closer agreement were primarily the curers, but this

same tendency was also found for women who were older, and presumably more experienced in dealing with illness. The research design and analytic process used here is broadly applicable to situations characterized by differential opportunities to learn about a cultural domain.

The research on cultural understandings about illness was part of a larger project on how families make treatment decisions among different alternatives. For Pichátaro, one of the findings was that accessibility constraints, such as lack of money or transportation, often contributed to decisions not to seek a physician's services, even though that was the treatment option preferred by the family. However, at the time of our research, an alternative explanation for low rates of physician utilization in communities like Pichátaro centered on the presence of ideas about illness that are incongruent with biomedical theory. So, the question arose whether changes in accessibility would indeed be associated with higher rates of physician utilization. We carried out a comparative study with a second Purépecha community, one that had better access to physician services. Using responses to two structured interview formats (including a term-frame substitution task) to carry out a number of different comparisons, we found no support for community-level differences in understandings about illness. Illness case histories collected over the same time period in both communities, however, showed differences in rates of resort to physician services, supporting our hypothesis that the higher rates of physician utilization in the comparison community reflected better access to physician services rather than a stronger biomedical orientation in illness knowledge (Young and Garro 1982).

I turn now to the Anishinaabe community which presents a somewhat more complicated perspective on the use of structured interviews in anthropological research. Much to my initial surprise, the "listing" part of the "free listing" interview fell somewhat flat. Typically, only a small number of illness names were mentioned before the individual claimed to be unable to recall any more. Overall, the information obtained did not result in what I considered to be sufficient material to construct a representative term-frame interview format. Apart from these interviews, I had followed or participated in a number of conversations about illness and knew that what came up in the free-listing interviews did not come close to approaching the range of illness related understandings in the community.

Later, on reflection, I came to at least a partial understanding of why the free listing interview did not proceed in the manner I originally expected. First, a noun-based free-listing task is at odds with the verb-based structure of the language known, and used, by most in the community, Anishinaabemowin. In Anishinaabemowin, there is an extremely rich descriptive vocabulary for conveying somatic experiences. In everyday talk, descriptions – which allow individuals to impart the particularities of their personal experience or what can be observed in others (e.g., children) – are more common than illness labels. Yet, the free list interview asked for names. Those that were given tended to be in English, a language which organizes reality in discrete chunks. Second, when someone is ill, assessing the nature of the problem condition does not depend so much on determining the best label but rather on assessing the likely causal possibilities, especially if there are grounds for suspecting that a physician's treatment may not be the sufficient or the most appropriate. Talk about a specific illness episode may suggest a likely cause while still being framed in such a way as to leave causal possibilities open. Remaining at the level of describing symptoms is one way of maintaining an openness to alternative, potentially viable, causal framings. In contrast, to refer to an illness by one of the labels seen as having been introduced by biomedical practitioners as in the case of high blood pressure or diabetes essentially sets a boundary around the range of causal possibilities and implicitly conveys the message that the illness is one that is appropriately treated by a physician.

It would, however, be incorrect to leave the impression that the free-listing interview provided little useful information. A number of explanatory frameworks were raised and I learned how concerned community members were about the relatively recent emergence of diabetes and high blood pressure

in the community. I also had my first conversations about *ondjine*, a term that is applied to illness or misfortune said to occur "for a reason" and which can be attributed to specific types of transgressions. There is also significant variability in the extent to which individuals deemed *ondjine* to be a credible account of illness and misfortune, alerting me to attend to social patterning along these lines.

These initial conversations about *ondjine* were the start of a long learning process that unfolded over the entire period of my fieldwork. The descriptions and case examples accompanying the free-listing task, along with other cases I soon learned about, touched on only the most common causes of *ondjine*. It did not take long, however, before I realized the need for further inquiry. Illness cases that did not quite fit with my emerging interpretation really stood out and led to further discussions and more rethinking on my part. Over time I learned that while most people knew about the common explanations for *ondjine*, others had richer, much more detailed understandings about the range of situations that could lead to *ondjine* and of the moral frameworks that underpinned assessments that illness or misfortune occurred "for a reason." My understanding of this quite complex, attributional framework emerged rather slowly and incrementally, mostly through the collection of ongoing illness case histories (see Garro 2000b). It was only near the close of my research that I felt I had gained enough understanding to even design a structured interview that asked about a range of possible causes (questions cast as a series of "yes-no" statements). Because I was using the structured interview as a way of double-checking what I had learned about *ondjine* from knowledgeable individuals, I purposefully selected a sample composed of individuals over the age of thirty who spoke Anishinaabemowin and who did not reject *ondjine* as a possible explanation for illness.

Further, during the free-listing interview almost everyone mentioned high blood pressure and diabetes, mentions that were almost always accompanied by statements expressing concern about the relatively recent emergence of these conditions in the community. For diabetes in particular, some expressed fears that everyone might have it at some future point. In many of these conversations, multiple, often quite diverse, explanations of what leads a person to develop diabetes were advanced. Although I had not originally intended to study understandings about high blood pressure and diabetes, these interviews (and the encouragement of staff at the community-run health center), led me to interview individuals diagnosed with diabetes and high blood pressure. While the studies on diabetes and high blood pressure were done separately, I used the same basic research design. Using a dual-interview format, participants were first asked a semi-structured series of open-ended questions modeled on Kleinman's (1978) explanatory model framework, followed by a structured interview consisting of a series of "yes-no" questions. As in the Pichátaro study, comments and reflections made in the free-listing interview and elsewhere (comments made about the specific illness condition as well as other conditions) were used to create a series of "yes-no" statements. As in earlier research, one benefit of using the structured interview format is that participants often made additional comments to support or clarify their response. In addition, the yes-no statements provide a check for knowledge not mentioned in the more open-ended portion of interview, which may occur through omission or when a particular statement does not correspond to the individual's personal experience, but does represent knowledge about diabetes or high blood pressure learned through interaction with others. A formal analysis of the "yes-no" questions measured the extent of sharing and nonsharing in responses (cultural consensus analysis; see Romney, Weller and Batchelder 1986).

In the analysis phase of all of the studies that I have done using this design, close attention is devoted to the interplay between the responses given in both the structured and the more open-ended interviews. In other publications, I demonstrate how relying on responses from both interviews contributes to a deeper and more finely nuanced analysis than could be achieved by using either one alone. In addition to exploring sharing and variability in cultural understandings about diabetes and

high blood pressure in a single community (Garro 1988, 1995), I have also employed this design in a comparative project involving three Anishinaabe communities (and 88 interviewees; see Garro 1996). In an even closer analysis for a single community, I look at patterned variability in responses and the social grounding of this variability in divergent life experiences (Garro 2000a).

Throughout this discussion, I have emphasized how qualitative structured interview methods can complement what is learned through other research approaches. There are many ways of integrating structured methods into the "ethnographic research cycle" – they can be used, for example, to explore and learn more about a domain, to assess a researcher's emerging understandings, to better understand sharing and variability, and to test hypotheses about the distribution of cultural knowledge. Given the time constraints present in any field project, the ethnographically appropriate use of structured data collection methods can be a boon to the field researcher. If ethnographically feasible, the benefits of collecting data that allow for direct comparisons across individuals should not be underestimated. Still, it must always be remembered that the reliable interpretation of results using structured data collection methods does not stand apart from other forms of evidence. And, as the discussion of the free-listing task in the Anishinaabe community and the abandonment of the plan to develop a term-frame interview illustrate, these tools themselves are based in assumptions about the nature of reality. There is no methodological recipe that will fit all settings. Yet, with reference to this same example, it should also be noted that the "ethnographic research cycle" was not impeded. Building on what was learned in the process, the research simply moved on to pose and address different ethnographic questions.

References

Agar, Michael H. (1996). *The Professional Stranger: An Informal Introduction to Ethnography,* 2nd edition. San Diego: Academic Press (first edition published in 1980).

Garro, Linda C. (1986). "Intracultural Variation in Folk Medical Knowledge: A Comparison Between Curers and Non-Curers." *American Anthropologist,* 88: 351-370.

Garro, Linda C. (1988). "Explaining High Blood Pressure: Variation in Knowledge about Illness." *American Ethnologist,* 15: 98-119.

Garro, Linda C. (1995). "Individual or Societal Responsibility? Explanations of Diabetes in an Anishinaabe (Ojibway) Community." *Social Science and Medicine,* 40: 37-46.

Garro, Linda C. (1996). "Intracultural Variation in Causal Accounts of Diabetes: A Comparison of Three Canadian Anishinaabe (Ojibway) Communities." *Culture, Medicine and Psychiatry,* 20: 381-420.

Garro, Linda C. (2000a). *Stirling Award Essay* "Remembering What One Knows and the Construction of the Past: A Comparison of Cultural Consensus Theory and Cultural Schema Theory." *Ethos,* 28: 275-319.

Garro, Linda C. (2000b). "Cultural Meaning, Explanations of Illness, and the Development of Comparative Frameworks." *Ethnology,* 39: 305-34.

Kleinman, Arthur. (1978). “Concepts and a Model for the Comparison of Medical Systems as Cultural Systems.” *Social Science and Medicine*, 12B: 85-93.

Romney, A.Kimball, Susan C. Weller, & William H. Batchelder. (1986). “Culture as Consensus: A Theory of Culture and Informant Accuracy.” *American Anthropologist,* 88: 313-338.

Spradley, James P. (1980). *Participant Observation*. New York: Holt, Rinehart and Winston.

Young, James C. & Linda C. Garro (1994 [1981]). *Medical Choice in a Mexican Village.* Prospect Heights, Illinois: Waveland.

Young, James C. & Linda C. Garro. (1982). “Variation in the Choice of Treatment in Two Mexican Communities.” *Social Science and Medicine,* 16: 1453-1465.

Standards of Rigor in Anthropology

Susan Weller
Department of Preventative Medicine and Community Health
University of Texas Medical Branch

The standards of rigor in anthropology are essentially the same as standards of scientific rigor across all of the social sciences. (This, of course, does not mean that all anthropologists adhere to these standards.) Although qualitative methods of measuring outcome variables can be used in experimental and evaluation research, these types of research designs are rarely used in anthropology. Anthropological research most often uses observational research designs (as opposed to true experiments with random assignment of subjects) with convenience sampling, and so I focus on those types of studies here. The most important standard is the statement of purpose (research question to be answered or hypothesis that is posed) and use of a research design that allows for the question to be answered, e.g., to contribute meaningful information and to allow for falsification of the hypothesis. A second issue concerns interview methods and the reliability of the data that are collected. A final issue concerns the coding and analysis of data. A key issue is recognizing the strength of qualitative methods, in contrast to more structured interviewing methods. Qualitative methods are most appropriate for areas where less is known about the topic.

With regard to the research design, researchers must specify the number of groups in the study (whether a single group or multiple groups are to be used), the rationale for selecting a group, and consideration for extraneous variables. If a project has only a single group of informants (without comparison groups) it is impossible to tell if themes mentioned during interviews are unique to that group or are simply general themes and would be mentioned by contrasting groups of individuals. For example, qualitative research with diabetic patients where diabetic patients are interviewed about their disease without a comparison group (either another group of patients with a chronic illness, a group of people who have no chronic illness, or even a comparison between subgroups of diabetic patients) can only find very general themes (see, for example: Quatromoni, Carballeira, Milbauer, Brunt, et al 1994; Anderson, Goddard, Garcia, et al 1998). In contrast, a study that, by design, interviews diabetic patients with good glycemic control and another group with poor glycemic control can identify themes that may be relevant to the management of diabetes (Savona, Miller, & Quandt 2004). Themes mentioned with similar frequency in both groups are of less interest than are themes mentioned more by one group than another. Also, extraneous factors can be controlled for by using matching procedures in the grouping design. Matching was used by Rubel et al (1984) to control for age and gender in his study of the folk illness *susto*. In the case of diabetes, in order to focus on lifestyle variables, those with good control and those with poor glycemic control should be matched for their duration of disease and their medications (De Alba et al., under review). Thus, when informants in the two groups are approximately similar with regard to the matching characteristics, the effects of those extraneous factors are removed, and factors of interest can be highlighted. A good study design is important whether a study uses qualitative or quantitative methods to assess the variables of interest.

A related issue with regard to research design is a clear description and rationale for informant selection. There are a few outstanding source books for this, especially with regard to qualitative research. The first is *The Ethnographic Interview* by Spradley (1979). Spradley is appropriate for research projects during initial or preliminary stages of interviewing. The focus is on how to get started, including who to interview and how to figure-out what questions to ask. The best single source on selecting informants is the Sage publication in the Qualitative Series by Jeffrey Johnson (1990),

*Selecting Ethnographic Informants***.** Johnson describes how groups of informants should be selected according to theoretical interests, e.g., the purpose of the study. For example, if ethnicity and feeding method are important to the study of infant feeding method preferences, then, informants should selected using these categories (e.g., breast-feeding Hispanics, bottle-feeding Hispanics, breastfeeding Anglos, and bottle-feeding Anglos) and there should be a sufficient number of informants in each of the categories (see Weller & Dungy 1986). Although there are many books on representative sampling, Johnson (1990) is the only one that discusses how to select a stratified convenience sample, a key part of most qualitative research.

Besides an appropriate research design and logical rationale for informant selection, studies need to use interviewing methods that obtain reliable and valid data. For some, one decision is whether to use group or individual interviews. Focus groups are very popular, but often researchers fail to understand that their effective sample size is somewhere between the number of groups and the number of participants, and is *not* the total number of participants. Evidence for reaching the saturation point and adequacy for the sample should be provided in descriptions of the work. Saturation is the point at which there are few or no new items or themes. Work by Fern (1982, also discussed in Morgan 1996) indicates that individual interviews are more efficient in eliciting topics; focus groups retrieve only about 60% of the amount of information as do individual interviews with the same number of informants. For highly shared domains, free-recall listing interviews (Weller & Romney 1988) can reach saturation with as few as 10 to 20 people. Similarly, focus groups discussing topics with high agreement across groups are reported to reach saturation with as few as 4-6 groups of about 8 people each. One thing is clear: the total number of hours spent in focused interviewing is directly related to the amount of information obtained. Two studies that used focus groups to interview Latino diabetic patients about their disease using four groups of approximately eight people each, had six to eight total hours of interviewing (Quatromoni, et al 1994; Anderson, et al 1998). In contrast, Hunt, Pugh, and Valenzuela (1998) conducted similar interviews individually with diabetic patients with slightly over 100 total hours of interviewing (51 patients, with approximately two hours per patient). The amount of information obtained by Hunt et al is evident in the series of papers published from those interviews.

The coding and analysis of data needs to include an assessment of the reliability of coding procedures. Some techniques use minimal interpretation of responses and use verbatim words or phrases (see free-recall listing, Weller & Romney 1988). Others interpret responses and form categories of responses. Methods that require interpretation need to include some method of inter-rater assessment of coding responses in to those categories. A standard way to handle such data is to specify the coding rules and criteria and then have at least two people independently code the text responses. Although this step is sometimes ignored, it is essential if the codes will be used in further analysis. Hruschka, Schwartz, St. John, et al (2004) describe an iterative procedure for bringing coders to acceptable levels of inter-rater agreement.

Cross-Discipline Emphasis

Qualitative interviewing methods can be used at any stage in a research project. Open-ended interviews can be used in the initial stages of a project, and the results can be used to design structured interview materials. Open-ended interviews can be conducted after a survey has been completed, to better understand responses to structured questions (Kempton, Boster, Hartley 1995; Baer 1996). Or, a study may rely exclusively on qualitative interviewing. The difference across the social sciences is mainly in the emphasis and where qualitative methods are integrated into the research process. Using broad generalizations, psychologists have "pilot" studies or an "instrument development" phase where open-ended questions are used to elicit themes for scale development. In psychology and sociology,

qualitative interviews are most often used during a "pilot" phase to explore how respondents interpret survey questions and what responses might be expected (Fowler 1995). Anthropology, in contrast, has relied upon qualitative methods (open-ended questions) for most studies, whether the methods are used alone or are integrated with quantitative methods. As a result, anthropology has developed considerable expertise in this area.

References

Anderson, R. M., C. E. Goddard, R. Garcia, J. R. Guzman, and F. Vazquez. (1998). "Using Focus Groups to Identify Diabetes Care and Education Issues for Latinos with Diabetes." *The Diabetes Educator,* 24: 618-625.

Baer, R. D. (1996). "Health and Mental Health Among Mexican-American Migrants: Implications for Survey Research. *Human Organization*, 55: 58-66.

Garcia De Alba Garcia, J. E., A. L. Rocha, I Lopez, R. D. Baer, W. Dressler, and S. C. Weller. "Diabetes is My Companion: Lifestyle and Self Management among Good and Poor Control Mexican Diabetic Patients." *Under Review*

Fern, E. F. (1982). "The Use of Focus Groups for Idea Generation: The Effects of Group Size, Acquaintanceship, and Moderator on Response Quantity and Quality. *Journal of Marketing Research*, 19: 1-13.

Fowler, F. J. (1995). *Survey Research Methods*. Applied Social Research Methods Series, Vol. 1, Newbury Park: Sage Pub.

Hunt, L. M., J. Pugh, and M. Valenzuela. (1998). "How Patients Adapt Diabetes Self-Care Recommendations in Everyday Life. *Journal of Family Practice*, 46: 207-215.

Hruschka,D. J., D. Schwartz, D. C. St. John, E. Picone-Decaro, R. A. Jenkins, and J. W. Carey. (2004). "Fixed-Choice and Open-Ended Response Formats: A Comparison from HIV Prevention Research." *Field Methods,* 16: 184-202.

Johnson, J. C. (1990), *Selecting Ethnographic Informants*. Qualitative Research Methods Series, Vol. 22. Newbury Park: Sage Pub.

Kempton, W. J., J. S. Boster, and J. A. Hartley. (1995). *Environmental Values in American Culture.* Cambridge: MIT Press.

Morgan, D. L. (1996). "Focus Groups." *Annual Review of Sociology*, 22: 129-52.

Quatromoni, P. A., N. P. Carballeira, M. Milbauer, M. Brunt, B. M. Posner, and S. R. Chipkin. (1994.) "Use of Focus Groups to Explore Nutrition Practices and Health Beliefs of Urban Caribbean Latinos with Diabetes. *Diabetes Care*, 17: 869-873.

Rubel, A. J., C. W. O'Nell, and R. Collado-Ardon. (1984). *Susto: A Folk Illness*. Berkeley: University of California Press.

Savoca, M. R., C. K. Miller, and S. A. Quandt. (2004). "Profiles of People With Type 2 Diabetes Mellitus: The Extremes of Glycemic control." *Social Science and Medicine,* 58: 2655-2666._

Spradley, J. P. (1979). *The Ethnographic Interview*. New York: Holt, Rinehart, and Winston.

Weller, S. C., and C. I. Dungy. (1986). “Personal Preferences and Ethnic Variations Among Anglo and Hispanic Breast- and Bottle-Feeders.” *Social Science and Medicine*, 23: 539-548.

Weller, S. C., A. K. Romney. (1988). *Systematic Data Collection*. Qualitative Research Methods Series, Vol. 10. Newbury Park: Sage Publication

Appendix 4.

Law and Social Science

Defining Excellence in Qualitative Research: Group Report for Law & Social Sciences

Defining the Field of Law and Social Sciences

The domain of Law and Social Sciences (LSS) is best defined simply as those areas of and approaches to research that fall within the purview of the so-named NSF Program. The domain includes work in cultural anthropology, sociology, criminology, law and economics, game-theoretic work in political science, and other fields. It excludes work that the NSF program directors consider to be outside of social sciences (which usually means in the humanities), including much historical work, critical theoretical approaches, and political theoretical work on law unless an empirical component is involved. LSS thus overlaps the fields constituted by the Law & Society Association (LSA) and the American Society of Criminology (ASC) and their journals but is not coterminous with them. For example, LSA includes work in the humanities but less of the explicitly rational-choice work funded under LSS.

The Methodologies of LSS

In a multidisciplinary field such as LSS, researchers have to ground their work in a particular tradition of research methodology and also make its value apparent to readers from other traditions. Of course, ideally all researchers do this, but it is especially important for work in a multidisciplinary field. The tradition may be one associated with a discipline, such as survey research or ethnography, or it may be best situated in an established body of literature within LSS, such as courtroom studies of plea bargaining or studies of the legitimacy of legal institutions, both of which have drawn from more than one discipline.

LSS realized from the start that the field goes beyond courts and other law-"looking" institutions to small-group interactions, and to the normative dimensions of other institutions. This decision has had methodological consequences, in that LSS has highlighted effective methods for studying legal processes and normative dimensions of social processes in a way that builds multi-disciplinary work and uses both quantitative and qualitative methods.

Qualitative methods have been particularly fruitful in LSS. In particular, the extended case method has proven to be exciting for illuminating broader social processes as well as analyzing the operations of courts and other law-like institutions. LSS researchers have built up traditions of interpreting judicial and other law-like reasoning processes, and have highlighted the interpretive character of law, the importance of studying micro-social processes, and especially the language in which social processes are carried out. For example, a common qualitative approach in studies of offending has been the use of in-depth interviews with active offenders, to better understand the situational dynamics associated with crime.

Standards for Researchers and Evaluators

Because of the complex nature of the field, the challenge to LSS scholars is producing proposals that are convincing across disciplinary and methodological boundaries. Proposals submitted to LSS (or mutatis mutandis to other multidisciplinary panels) are read by reviewers from more than one discipline (This situation is not unique to the NSF; the same is true with many foundation research review panels, for example, the Social Science Research Council and the National Endowment for the Humanities). Successful principal investigators (PIs) write proposals that can make sense to a varied set of readers. What is involved in "making sense"?

Project Framing

In describing the overall purpose of their proposed project, PIs should:

- **Articulate a clear research question.** In some cases it may be appropriate to frame questions in terms of hypotheses, but often it may not be. Science can mean pursuing questions in a logical and systematic way, and may be inductive in its orientation. Moreover, PIs should express clearly how law is central to their research question.
- **Describe why the research is important**, what the research product might look like, and why the result will be exciting. The proposal should convince reviewers across fields that the findings of the project will have intellectual, political or social significance.
- **Appropriate use of theory.** Projects might be oriented toward advancing, developing or refining theory in a particular field, or studies might be shaped by, or intended to speak to, a particular body of theoretical literature, but then they must specify what "theory" means and how their project stands vis-à-vis a particular body of literature. Moreover, PIs must be aware of how "theory" takes on different meanings in different disciplines.
- **Use language that is convincing across disciplines.** Moreover, in order to persuade readers who may have only rudimentary familiarity with the concerns and traditions of her or his disciplines, the researcher should specify, define, and operationalize terms and avoid field-specific jargon.

Research Design and Methodology

The research design and methodology portion of LSS proposals should describe clearly case selection, sampling procedures, data collection plans, and data analysis schemes. PIs should take particular care to:

- **Justify case selection and sampling procedures.** In qualitative research, the idea of "sample" is important but takes on different forms depending on the topic being studied. For example, a PI may study recurring processes in a particular institution, such as a court or a hospital, and these many events constitute a set of distinct data points. Since this procedure may deviate from some reviewers' understanding of a "sample," the PI needs to make explicit the relationship of this institution to a broader set of such institutions. Cases may be selected for a variety of reasons. Some research may require triangulating using different sites in order to make broader claims about types of institutions or processes; other projects may emphasize contrasts and differences. Or, certain processes may be particularly evident in some institutions, researchers may be interested in emergent processes or deviant cases or specific dilemmas. However, PIs must make these selection criteria explicit.
- **Display a dialectic between research methodologies and field contingencies.** Researchers should show that they have not only thought through the steps of their work but also anticipated and made explicit how they would respond to contingencies ("what if X does not work out?"). Qualitative research inevitably involves a dialectic between research methodologies and field contingencies.

The NSF's Role in Promoting Qualitative Research

Increased Funding for Pilot Research

It is easier for applicants to demonstrate both their rigorous thinking through of method and demonstrating their awareness of and potential responses to contingencies if they can undertake preliminary research. Funding preliminary field research is critical to aid qualitative researchers in working out these contingencies. Pre-dissertation field grants might have an option of methods training, but such training should not be required. For the same reasons, pilot grants, perhaps in partnership with the appropriate professional associations (an expanded version of SGER—Small Grant for Exploratory Research funding opportunity), with no or minimal overhead, would allow researchers at all levels to prepare better-quality proposals.

Enhanced Resources for Teaching Qualitative Methods

In addition to increased funding for preliminary field research, workshops for teachers of qualitative methods, perhaps in partnership with Law & Society Association or the American Society of Criminology, would bring to many of their members with legal and qualitative backgrounds a better competence at designing NSF-fundable research.

Fostering Links Between Qualitative Researchers

Moreover, the NSF should foster collaborative, comparative work amongst qualitative researchers. Triangulating and selecting multiple sites to bring out contrasts makes for more valid research, but often requires collaboration among several scholars, which requires more time and more resources. The NSF could facilitate such collaborative projects through providing increased funding for comparative work. Finally, LSS produces a great deal of data that could be analyzed by other researchers. The NSF should explore ways to partner with scholarly associations to preserve and make available different kinds of written or recorded information.

Law and Social Science
Papers Presented by Participants

Warranted Trust

Don Brenneis
Department of Anthropology
University of California, Santa Cruz

Thinking about questions of rigor, imagination, and scientific value within the context of Law and Social Sciences is a bracing experience, in part because the intellectual ambit is defined by core topical concerns rather than by specific disciplinary programs or practices. The rubric extends across a wide range of theoretical and methodological perspectives, and the experiences of those of us who have participated in various of the law and social sciences panels speak directly to issues of translation, clarity, and disciplinary divergence (and convergence) central to this workshop. I am also, nonetheless, necessarily writing as an ethnographer, and more specifically as an ethnographer with a particular commitment to the principled examination of language and interaction in multiple contexts, so several of my comments below are likely to be supplementary to those articulated by my anthropological colleagues. I also am writing as a former editor of a major journal in sociocultural anthropology and so recurrently have been afforded the chance to think very concretely about questions of rigor in ethnographic writing.

As a framing comment, I would like to suggest that the assessment of rigor is not a global judgment, and that we necessarily have different expectations and turn to different criteria and kinds of criteria at various stages over the course of a research project. More concretely, there are at least three different moments in the life of a specific research project when we might be involved as evaluators: as proposal reviewers (whether ad hoc or as panel members), as manuscript reviewers (whether ad hoc or as editors), and as readers who draw upon the published work of others. Somewhat different understandings of rigor and of what can be taken as evidence of it come into play at each of these moments: when we are asked to consider a proposed trajectory (in the future tense), when we are both evaluating and helping guide the ways in which arguments and presentations based on work accomplished, and when we are evaluating the status and value of published (or otherwise disseminated) research with an eye to how it might shape our own science and scholarship. And these differences, especially between reviewing proposals and manuscripts, might well be even greater when we consider qualitative research, as a frequent hallmark of such work is the flexibility it affords in addressing unanticipated but empirically significant discoveries when in the field.

I will try here to be both candid and brief about what I looked for as evidence of rigor in proposals and in manuscripts. As a proposal reviewer, especially on as far-reaching a panel as that in law and social sciences, I often found myself reading proposals concerning topics and proposing methodologies with which I was relatively unfamiliar. I necessarily placed a great deal of weight on the logic of the research plan, which depended in turn on the clarity of the methodologies proposed and the explicit links made between specific research questions and strategies for pursuing them. As an ethnographer, I also looked for evidence that the proposed questions and methods made sense in and for the context in which the research was to be pursued, which in turn depended upon the proposer locating the work effectively, if economically, vis-à-vis the specific site(s) of inquiry. Beyond rigor, I also looked for a strong and imaginative sense of research problem (and of ways of pursuing it), for indications that the proposer had the capacity to reshape the research in principled ways should their data surprise them, and for underlying ideas that both articulated with and moved forwarded theoretical questions. I should also note here that I have served for a number of years on Anthropology panels at the National Endowment for the Humanities; in evaluating proposed research, we focused on exactly the

same criteria. Proposals are necessarily in the future tense, and reviewers necessarily look for evidence of the capacity, logic, imagination, and senses of principled possibility that would warrant our trust in the researcher and the research.

As a manuscript reviewer and editor, on the other hand, one is reviewing work that has already been completed, and a great deal of the evaluative weight rests on the relationships among method, data, and argument as presented in the text. When thinking about rigor, my own standards for manuscript evaluation include such dimensions as the effectiveness of contextualization, the relevance of the variables being pursued to the questions at hand, the internal analytical consistency of authors' consideration of their data, explicitness in laying out both the argument and the data, and provision of sufficient descriptive materials to allow readers to judge the effectiveness of the proposed analysis. Rigor by itself, obviously, is never enough to warrant publication; it must be put to the service of furthering, refining, and transforming our understandings of the phenomena we study.

I want here to note some of the kinds of qualitative methods, perspectives, and opportunities that figure in ethnographic research, whether in sociolegal studies or more generally. The list is meant to be suggestive rather than exhaustive, and my comments are quite abbreviated:

- Structured interviewing can obviously be both rigorous and powerful; it has long been a central strategy in qualitative research, for good reason. At the same time, two caveats seem important. First, this strategy most likely works best when pursued in complement to other types of qualitative work. Second, questions of the context-appropriateness of interviewing are critical. While as social scientists we might view interviewing as neutral and unexceptionable, many of our consultants might understand it quite differently, as Briggs (1986) has argued.
- Other, perhaps locally-defined genres of conversation, instruction, and the like, ones that are less susceptible to a priori topical and sequential structuring but may be quite revealing. In some of these genres the researcher may be an active participant, in others an open-eared observer.
- An ongoing concern, the forms of which take different shapes for different ends and in different sites, for observation, situating research and local practice, and the like.
- Narrative elicitation and analysis, especially life histories. A particularly provocative and convincing example of such work in sociolegal studies is Engel and Munger's (2003) study of the consequences, legal, social, and personal, of the Americans with Disabilities Act, one that affords understandings that a more orthodox consideration limited to subsequent legal cases and appeals could not offer.
- Discourse analysis in the "small d," linguist's sense of discourse, i.e., taping, transcribing, and analyzing naturally occurring, connected speech in situ. Some fine examples of the detailed analysis of conflict discourse can be found in Watson-Gegeo and White (eds.: 1990) and Briggs (ed.: 1996). A further elegant and powerful example of such work in the sociolegal arena is Mertz's (2007) study of pedagogical talk (and, most specifically, of instantiations of the Socratic method) in first year law school classrooms. Talk, whether in a classroom or in the context of village disputes, is both about social practice and social practice in itself. Thorough, systematic, and rigorous analysis of talk can effectively illuminate both its subjects and its social consequences and implications.
- As Mertz's work and a recent book by Amsterdam and Bruner (2000) brilliantly demonstrate, careful and systematic attention to the contexts and practices of socialization – of socialization into profession, craft, or other kinds of sociocultural knowledge, practice, and position – can provide powerful social and cultural data.

- Various kinds of case-based approaches to the course of related events over time have long figured centrally in both social anthropology and sociolegal studies (e.g., in the work of the Manchester School). Such cases not only provide an approach to understanding specific disputes and the like; they also often shed light on much broader materials for social analysis. Rigor of a more conventional kind is often easier to pursue in a specific synchronic moment; systematic and comparative case-based research can help provide a rigorous basis for the usually less tractable consideration of related social activity over time. And here the use of the case method is oddly parallel to the examination of discourse, as in both examples it is the connectedness of action over time, whether within the short-term frame of a specific verbal interaction or across months or years.

I want to address two other issues quickly here. The first concerns the divergence issue, where variation often has less to do with methodology per se than with the polysemy of "theory" as a term. As a quick example, I found during my time as editor of *American Ethnologist* that manuscripts drawing heavily upon either quantitative data or formal modeling often elicited responses from reviewers characterizing them as "undertheorized," while colleagues working in such quantitative traditions found the journal's articles frequently interesting and engaging but argued that they generally did not contribute to "theory building." How we define the goal of our inquiry and the nature of better and less effective epistemological means to those goals are critical questions. We often, in not examining our own preconceptions carefully enough, fall into false dichotomies.

A final point has to do with what I see as a potentially invaluable resource for considering the rigor of published work. Making our data public, i.e., providing for subsequent readers at least some of the materials on which we base our analysis – or at least our own first-pass renditions of such materials as cases, transcripts, and the like - has long been desirable and is now, with the possibility of electronic archiving and distribution, closer to feasible. Replication of social science research in the classic sense is rarely possible, but the reanalysis of relatively raw material, or at the least the more thoroughgoing reading of publications with an eye to the materials on which they are based, now might be more likely. Linguistic anthropologists have long been frustrated by the inability of most journals to publish extensive transcript material, just as many doing sociolegal research have found it impossible to publish the extended cases upon which they base their interpretation. With the increasing availability of electronic data bases (an example being the development of the American Anthroplogical Association's AnthroSource), a very inexpensive resource for such storage, access, and, most important, reanalysis could become broadly available. NSF might want to consider hosting such extended empirical materials across the social sciences.

References

Amsterdam, Anthony, and Jerome Bruner. (2000). *Minding the Law*. Cambridge: Harvard University Press.

Briggs, Charles L. (1986). *Learning How to Ask: a Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge: Cambridge University Press.

Briggs, Charles L. (1996) *Disorderly Discourse: Narrative, Conflict, and Inequality*. New York: Oxford University Press.

Engel, David M., and Frank W. Munger. (2003). *Rights of Inclusion: Law and Identity in the Life Stories of Americans with Disabilities.* Chicago: University of Chicago Press.

Mertz, Elizabeth. (2007). *The Language of Law School: Learning to "Think Like a Lawyer."* New York: Oxford University Press. *Forthcoming.*

Watson-Gegeo, Karen Ann, and Geoffrey White. (1990). *Disentangling: Conflict Discourse in Pacific Societies.* Stanford: Stanford University Press.

Qualitative Research in Law and Social Sciences

Susan Bibler Coutin
Department of Criminology, Law and Society
University of California, Irvine

Sociolegal scholarship employs a wide range of qualitative research methods, including ethnographic approaches, interviews, observations, focus groups, textual analysis, archival work, and more. In my experience, researchers who are unfamiliar with qualitative studies sometimes misunderstand what such methods are designed to accomplish. Furthermore, even among researchers who use qualitative methods, there is sometimes a tendency to treat ethnography as either a black box or part of a tool kit, both of which are problematic. I will discuss these issues before describing promising areas of qualitative research.

Both quantitative and qualitative research use empirical methods to decipher the workings of social, cultural, and legal processes. They differ, however, in how they go about this deciphering. Quantitative research is designed to test hypotheses. Factors to be considered in assessing quantitative research designs include external validity (does a proxy really approximate what it is supposed to?), the construction of the sample, the presence of confounding factors, the appropriateness of the pool from which the sample is drawn, selection effects that may arise in forming a sample, the generalizability of the findings, the falsifiability of the hypothesis to be tested, and the replicability of the study. These factors center on whether or not the "test" conditions -- whether experimental, a survey, or an aggregate data set -- accurately mirror broader reality. A quantitative analysis will only produce valid results if the data are of high quality in the first place.

Qualitative research often attempts to answer a question rather than to test a hypothesis. Instead of devising "test conditions," qualitative researchers examine on-going social processes, study records or artifacts that shape or are produced by these processes, and talk to people who are engaged in or affected by the processes being studied. Issues to be considered when assessing qualitative research designs include the following: Is the case uniquely appropriate for the study in question? Can the research question be answered with the sort of data that is to be gathered? Is there a broader phenomenon that is being studied through a particular case? Will the study advance theory regarding this phenomenon in particular ways? Is the research question relatively open; that is, are there multiple possible answers to the question? Is it clear that the researcher needs the data in order to answer the research question? Will the data collection be thorough? Will the research process give the researcher the expertise that will make it possible to answer the research question? These issues focus less on the quality of the data per se, and more on the appropriateness of the case and the quality of collection. A high quality study will produce rich and complex (multi-faceted) knowledge of particular phenomena.

Unfortunately, researchers sometimes treat "ethnography" in particular as a self-explanatory process or as part of a tool kit that can be casually deployed. First, it is not clear that "ethnography" is a single method rather than a type of account or a way of perceiving and representing social and cultural (including legal) phenomena. What makes a project "ethnographic" may be this mode of perception and representation rather than a particular set of activities (e.g., conducting observations, doing interviews). It is possible to produce an ethnographic account without doing fieldwork, as traditionally understood. And it is possible to use "ethnographic methods" but to produce an account that might not be recognized as an ethnography. For instance, researchers may work with qualitative data, but analyze them in a

quantitative fashion, such as through coding data in order to identify representational trends, but not investigating how such representations are produced, read, or interpreted. Second, "ethnography" is sometimes tacked onto a list of research methods, as in "and I will do an ethnography of X," without an explanation of what such an endeavor entails. Ethnography becomes a black box, a mysterious process, opaque because it is seemingly transparent. Third, ethnography is sometimes treated as a research technique that can be deployed by a researcher at will. In fact, ethnographic research may require forming relationships, becoming part of a research setting, acquiring new forms of expertise, and developing (or suspending) particular ways of seeing. Ethnographic methods can be employed by scholars in numerous disciplines and can be combined with other approaches, but ethnographic research is often only possible with some collaboration from ethnographic subjects (and therefore is not deployed at the sole discretion of a researcher) and can entail the complex (and potentially compromising) positionings.

To avoid the "black box" and "tool kit" problems, scholars who are proposing qualitative studies (whether or not these are ethnographic) can elaborate on what they plan to do and why they plan to do it. It is helpful to work up the case(s) to be studied. What unique opportunities does this case present? How do the theoretical issues to be addressed play out within this particular case? If observations are to be conducted, how will these allow the researcher to learn about situations and events that will shed light on the research question posed? For example, to identify the normative orders that influence police control of territory, Steve Herbert (1997), a geographer, went on ride-alongs and observed officers' exercise of discretion. Similarly, to understand the linguistic strategies of Muslim women who had to testify in a traditionally masculine legal setting, Susan Hirsch (1998), an anthropologist, observed and recorded hearings in an Islamic court in Kenya. If interviews are to be conducted, who will be interviewed, about what topics, and how will the interview data speak to the research question? In my own discipline of anthropology, interviews (and observations) are generally conducted by the researcher (rather than delegated to a research assistant or outsourced to a company) and in a language that both the researcher and the interviewees speak fluently. There is supposed to be an "immediacy" to the research process. Qualitative methods can include the analysis of texts, and when this is the case, it is helpful to explain the sorts of insights that are available from particular sources. For example, to analyze ways that low-level immigration officials understood and applied the Chinese Exclusion Act in the late 1800s, Kitty Calavita (2000), a sociologist, studied correspondence between these officials and their supervisors regarding the correct response in 'problematic' or ambiguous cases. Qualitative studies now frequently employ a computer program to code and sort data, however, indicating that such a program is to be used does not adequately describe how data is to be analyzed. Instead of simply mentioning a program, it is helpful for researchers to explain how they are going to go about making sense of their data -- what sorts of things will they be looking for or paying attention to within documents, fieldnotes, interview transcripts, and other sorts of material? Why are these particular facets of accounts significant? To what will their data be speaking, and how will researchers discern what is being said?

It is difficult to specify particularly promising topics for qualitative study in law and social sciences, as there are so many possibilities. I will therefore, of necessity, be idiosyncratic, and describe some of the topics that I find most interesting. First, a recent convergence between scholars in science and technology studies, law and social sciences, anthropology, and even information technologies directs attention to forms of assessment -- essentially, what we are engaged in here today -- and particularly to ways that law and other forms of knowledge (including the social sciences) derive from similar (even identical?) understandings of truth. Note, for example, the degree to which law and the social sciences use an overlapping vocabulary: evidence, validity, consent, case, evaluation, judging, credibility,

identity, proof, bias, to list a few significant terms. How are legal and other truth claims produced and assessed? What gaps or disjunctures arise in such processes? How do assessments shift over time, such that claims, activities, and persons that are considered illegitimate are sometimes redefined at subsequent moments -- and vice versa?

Second, qualitative methods may be particularly appropriate for analyzing powerful institutions that produce law and/or quasi-legal agreements, norms, and policies. I have in mind the United Nations, the World Bank, the International Monetary Fund, the European Union, USAID, and other such bodies. Qualitative methods, which have often been used to study particular communities, delimited sites (e.g., a street corner), or bounded groups are also well-suited for identifying reconfigured understandings of persons, states, rights, development, international relations, gender, and law that are being produced through such institutions.

Third, while doing a literature review for a paper on the remittances that migrants send to relatives in their countries of origin, I was struck by the hidden "qualitative" nature of certain quantitative studies. For example, a 2003 IMF Working Paper by Ralph Chami, Connel Fullenkamp, and Samir Jahjah used IMF and World Bank remittance data on 113 countries from 1970 to 1998 in order to determine whether immigrant remittances were a source of capital for economic development (they concluded that by and large they are not). In order to perform a quantitative analysis, the authors developed a model that reads much like a fictional account. I quote from their paper: "We envision a country made up of a large number of identical two-person families in which one of the members has migrated and is earning an exogenous income in the foreign country.... We refer to this person as the immigrant.... The family member who remains in the home country, whom we refer to as the recipient ... works in the domestic labor market. The recipient is risk averse and works for a risk-neutral firm. Output *x* in the recipient's country is uncertain" (p. 11). While the authors develop this model -- only a small bit of which I've quoted here -- in order to test relationships between remitting, risk, productivity, and economic growth, it occurs to me that this model is revealing as a piece of qualitative data. This model assumes that immigrants will feel resentful if recipients reduce their labor output due to receiving funds, that remitters are largely altruistic, and that recipients may be able to exploit remitters. In this case, scientific rigor-- a sophisticated analysis of whether relationships between variables were as predicted by the model -- required producing a fictitious "vision" of migration, countries, families, and work. In line with my call for research regarding assessment, it would be interesting to study ways that, whether qualitative or quantitative, scientific analysis (much like legal analysis), sometimes requires at least temporarily considering particular "fictions" to be valid.

References:

Calavita, Kitty. 2000. "The Paradoxes of Race, Class, Identity and 'Passing': Enforcing the Chinese Exclusion Acts, 1882-1910." *Law and Social Inquiry* 25:1-40.

Chami, Ralph, Connel Fullenkamp, and Samir Jahjah. 2003. "Are Immigrant Remittance Flows a Source of Capital for Development?" IMF Working Paper WP/03/189, International Monetary Fund.

Herbert, Steve. 1997. *Policing Space: Territoriality and the Los Angeles Police Department.* Minneapolis: University of Minnesota Press.

Susan F. Hirsch. 1998. *Pronouncing and Persevering: Gender and the Discourses of Disputing in an African Islamic Court.* Chicago: University of Chicago Press.

Zones of Methodological Convergence in Qualitative Social Science Research

John Bowen
Department of Anthropology
Department of Religious Studies
Washington University

A number of "middle-range methods" offer bridges across the several research traditions. These are strategies, ways of knowing something, less than theories but more than techniques; indeed, they each allow for a range of techniques.

- **Triangulation**-- a kind of sampling with a very small n, especially appropriate for studying complex institutions. We should explore whether it differs from the small n macro-sociological studies: probably the two do differ in terms of the selection of cases. Triangulation implies a "most different case" strategy. The use of controls to isolate causal variables is logically related to this.

- **Contrasts**-- studied for their own sake, not as a way to home in on something (as in #1). "Comparative causal analysis" might fit here. The point is to point up important differences across countries, or social features (urban/rural), for example. This approach might lead to something like # 1 but often does not.

- **Reliability/validity**-- a framework for evaluating methodological approaches across research traditions. The two are alternative "ways in" to judging how well we are approximating something going on in society. Anthropologists approach this desiderata through validity—long-term acquaintance with people, meanings, practices—while psychologists start with reliability—cross-rater, or through repeat experiments. But the idea is the same.

- **Case studies**-- they have the value of being naturally-occurring processes, and can be a shared focus for researchers from a number of traditions. Experiments are cases that occur in artificial environments; events observed in field situations are uncontrolled and natural processes. Perhaps the odd phrase "quasi-experimental" from Campbell holds promise for us.

Thinking about Standards in Qualitative Research

Wendy Espeland
Department of Sociology
Northwestern University

Our charge for this essay was to suggest standards of rigor that might be used in designing or evaluating qualitative methods and to discuss promising topics for research that use qualitative methods. Because there are many strategies for promoting rigorous research, especially in an inter-disciplinary field such as law and social science, I will focus on two: theory and multiple methods.

Theoretical Framing

One source of rigor in qualitative (and quantitative research) is the theoretical framing of the research. Most generally, theory organizes strategies for discovering similarities and differences in social relationships. Theory can be used to derive questions, explain why something is a puzzle, identify gaps in our knowledge, and frame expectations. Theory can also help explain what this particular research will contribute to knowledge by suggesting how the world has changed in ways that our theories cannot yet account for, by identifying patterns in one context and suggesting why these are modified or missing in another context, or by disclosing how this research will help us refine concepts or the scope of theory more precisely.

Among other things, theory is an orienting device that highlights or problematizes some parts of the world while ignoring others. This role needs to be made explicit and defended in research. Researchers should offer a convincing account of why this location, these comparisons, this event, or these people are a good place to study the questions or issues that motivate the research. An important part of rigorous theorizing involves describing what abstract concepts or relationships might look like on the ground, in the particular context one is studying them, with the kind of evidence one is mobilizing. So, for example, in her investigation of persons seeking legal redress from domestic violence in Trinidad, Lazarus-Black (2001) describes how "agency," a concept usually associated with personal efficacy, is negotiated in various encounters with police, clerks, or attorneys who may or may not serve a summons promptly, process an application appropriately, decide to adjourn a case and so on. In doing so, she makes explicit how "agency" is enacted or operationalized in her research, as well as making a convincing argument for why we need to reassess the meaning of the concept. In addition to spelling out the connections between key concepts and evidence, researchers should also, if possible, offer examples of evidence that might challenge expectations, call into question an argument or an interpretation. If a powerful group does not become involved in negotiating important legislation, for example, a researcher may need to spend more time analyzing how interests are defined in relation to this issue or focus more implementation than enactment.

In my account, theory does a lot of heavy lifting. It offers a framework for identifying questions, helps suggest locations for investigating those questions, and it offers a means for both specifying the particulars of a social location and for generalizing from particular contexts or cases to others. Theory also provides a framework for helping to elaborate the terms under which explanations are suspect or discredited, by suggesting the conditions or findings that would cause us to revise our predictions or reject our explanations. Making the link between theory and evidence as explicit as possible is a standard that can be extended to many styles of research. And doing so does not mean that a theoretical framing, or the methods that devolve from them, cannot be revised as the researcher's engagement with

the world deepens. One of the great virtues of qualitative methods, especially field work, is its flexibility which permits researchers to adapt to what they are learning, to follow their noses. Efforts to be precise in one's theoretically framing allow the researcher to see more clearly and quickly where and how expectations are wrong, or where concepts do not materialize in the forms one imagined, something which facilitates a deeply sustained, iterative relationship between theory and evidence.

Multiple Methods or Triangulation

Another strategy for promoting rigorous research involves incorporating different types of evidence collected in different ways, what some call "multiple methods" or "triangulation." My own research has mostly been conducted on or in organizations and I routinely use different strategies to gather evidence. The methods I have used most often include field work, where my role has ranged from full participant to solely an observer (although one is never purely anything in field work), interviews, both formal and informal, the analysis of contemporary and historical documents (e.g. newspaper stories, official documents as well as informal, private documents, and archived, historical sources of various kinds--letters, drafts of documents, memos). I have also on occasion conducted brief surveys of people who attend functions I am observing. And my collaborators have performed statistical analyses on quantitative data. Each of these methods calls for distinctive forms of expertise and skepticism.

The standard line on "triangulation" is that the strengths and limitations of different sorts of evidence can complement one another when sources are made to speak to each other. I believe this. So, for example, in studying the effects of law school rankings I am reassured when a statistical analysis of the applications and yields of law school applicants confirm what administrators and law applicants tell me: that a change in ranking affects people's decisions about where to apply and attend law school. (Espeland and Sauder 2007) The statistical analysis shows that the patterns my respondents describe extends to many different law schools and it helps me to be more precise in determining how big an effect a change in rankings has on people's decisions. But what the statistical analysis can't tell me, and what my respondents and observations can, is how this process works, the nature of the dynamic relationship among variables, or what Robert Merton (1968: 43-44) calls the "mechanisms" of social relations. And statistical analysis cannot say much about the meaning and construction of numbers that are the object of analysis, and the authority that members grant them, or not.

As my example suggests, there are many versions and uses of "triangulation." Different kinds and sources of data might be used to approximate the type of fact -checking or corroboration that journalists deploy to verify information. Mitchell Duneier's work (1999: 345-47) demonstrates both the labor and the value at stake in "checking stuff." Or different methods may be used to get at different but related questions, in a sort of division of labor. We may, for example, compare historical records with contemporary accounts to try to assess changes over time, or conduct intensive field work in one part of an organization and conduct interviews or review documents to learn about other parts. This division of labor may be organized to exploit the relative advantages of different approaches, it may be an expedient use of scare resources, or it may reflect the power relations of those we study, since the terms of access that researchers are granted to the people and places they study are saturated with power. This is one reason why, as many ethnographers have noted, people's reactions to a researcher's efforts to obtain access becomes crucial data to analyze.

Triangulation can take the form of collecting different kinds of evidence, all of which is directed toward answering the same questions or examining the same phenomena. In practice, this form of triangulation is of often organized as a series of comparisons. We compare the evidence collected from different sources in order to better understand the biases or omissions of each and to produce a more comprehensive view of the social phenomena we investigate. So, for example, we might compare

official accounts with informal ones, what people say --in interviews, surveys, informal conversations, emails or written reports, on their websites, to the media-- with what they do, what they do in one context with what they do in another, what they do over time, and so on. These sorts of comparisons will almost inevitably lead to a more complex account of what is going.

But it is not always easy to reconcile or even interpret the differences that such comparisons might reveal. Are these differences inconsistencies? Do they represent alternative views that correspond to different social locations? Do they reflect duplicity, politics, or people's sophisticated understandings of the constraints of genre or the demands of audience? There is no simply set of principles which allow a researcher to interpret the patterns or gaps that emerge in comparisons of these sorts but an effort to explain them will almost always prove fruitful. Responding to this sort of variation is one way of organizing the iterative relationship of research as it unfolds. I should add the field of law and society was founded in response to exactly this type of comparison. The discrepancy between what has become known as law on the books, the formal, textual law of legislation and judicial opinions that is the province of exegesis, and the unruly law as practiced, gave rise to a vibrant, interdisciplinary community that has spent almost 50 years trying to sort out the impressive variety of forms of expression that this gap can take.

Promising Areas of Investigation Using Qualitative Methods

As Susan Bibler Coutin points out in her essay (page 63), in an interdisciplinary field like law and social science, there are so many fruitful topics for qualitative research it is hard to limit them. Rather than naming particular topics, I want to encourage an approach to studying law that conceptualizes law as a form of authority in relation to other forms of authority. To emphasize law as one type of authority among many, scholars need to analyze how legal authority is claimed and sustained and how it reinforces, stymies, or mediates authority that might be grounded in culture, science, religion, kinship, professions, nationalism or whatever. How does law and its effects change as legal authority intermingles with other forms? This question begets others: why and when do non-legal forms of authority triumph? How do different groups appropriate legal authority differently? And what do we mean by law? Carol Heimer and Lisa Staffen's (1998) study of how social responsibility is organized in neo-natal intensive care units, and Susan Silbey and Patricia Ewick's (2003) analysis of safety regimes scientific laboratories, are two analyses that exemplify the value of this approach. Both show how collaboration and conflict among different forms of authority profoundly shape how members understand what law is and what law can do. If we do not investigate law in relation to other forms of authority we will fail to appreciate the full range of the effects of law, the complexity of legal authority, and the extent to which the legitimacy of law, however naturalized for some, is never permanently secure. This approach to studying law is well-suited to qualitative methods since relations of authority are often emergent, subject to on-going negotiation and interpretation, and difficult to reduce to variables.

References

Duneier, Mitchell. (1999). *Sidewalk.* New York: Farrar, Straus and Giroux.

Espeland, Wendy Nelson and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." American Journal of Sociology 113: 1-40.

Heimer, Carol and Lisa Staffen. (1998). *For the Sake of the Children*. Chicago: University of Chicago Press.

Lazarus-Black, Mindie. (2001). "Law and the Pragmatics of Inclusion: Governing Domestic Violence in Trinidad and Tobago." *American Ethnologist*, 28: 388-416.

Merton, Robert. (1968). *Social Theory and Social Structure.* New York: Free Press.

Silbey, Susan and Patricia Ewick. (2003). "The Architecture of Authority: The Place of Law in the Space of Science," in *The Place of Law*, edited by A. Sarat, L. Douglas, and M. Umphrey. Ann Arbor: University of Michigan.

The Status of Qualitative Research in Criminology

Jody Miller
Department of Criminology & Criminal Justice
University of Missouri – St. Louis

Standards of Rigor in Criminology and the Place of Qualitative Scholarship

Criminology's emergence as an independent discipline is fairly recent. In the U.S., criminology developed primarily as a subfield of sociology, and leading criminologists are now housed both in sociology and criminology programs. Criminologists have long felt marginalized within the discipline of sociology (with evidence of good reason), and this has been one of the factors contributing to its independent development. This split is significant, both in terms of situating the field in the historical linkages between the two disciplines, and in understanding the consequences of this division for the scholarly study of crime generally, and the place of qualitative research in criminology specifically.

Qualitative research was very much at the heart of early studies of crime. The field research tradition of the Chicago School – and its use of the urban landscape as a social laboratory – resulted in numerous studies of crime and deviance (see Adler & Adler 1987). As this model of research was contested within the broader field of sociology, so also it fell out of favor among those studying crime, particularly with the evolution of survey research methods and the advancement of statistical techniques (see Hagan & McCarthy 1997: 3-4).

While there remains a divide within sociology between qualitative and quantitative methodologies, and interpretive and positivist epistemologies, it is even more palpable within the discipline of criminology.[5] While criminology remains influenced by its sociological roots, its emergence as a separate discipline (with separate journals and separate audiences) has resulted in greater insulation of the discipline from theoretical developments in other social sciences.[6] This has limited the cross-fertilization of ideas across disciplines. It has also meant that many of the trends that have taken hold, or at least shaken up sociology's claim as positivist science – for instance, the postmodern and poststructuralist turns – have had minimal impact on the criminological enterprise. Moreover, because large scale crime research is funded by government agencies in justice and public health, often with the expectation of direct policy relevance, the claim to "scientific" rigor is ever important, and remains largely defined in quantitative and positivistic terms. This trend is reflected in the dominance of theory testing models, often with the use of large, complex datasets that require advanced statistical techniques.

Largely as a result of the revitalization of neighborhood studies of crime, as well as growing interest in situational aspects of offending, there has been a recent resurgence of appreciation[7] for qualitative research in criminology. Before discussing this further, however, I would first like to focus on some of the consequences of the positivist/quantitative dominance in the field for qualitative scholarship on crime.

First, the distinction between qualitative data collection methods and qualitative analysis is an important one. It has been common in the field to employ qualitative data collection methods or use qualitative data for quantitative analyses. This is evident in several areas of research. Policing scholars, for instance, have relied on large-scale observational studies of police/citizen interactions to analyze police decision-making (focusing on the impact of citizen behavior on arrest decisions, and more

5 My discussion here applies specifically to the U.S. context of criminology. In Europe, criminology is more closely linked with the disciplines of law, philosophy, and psychiatry.

6 This may result from examining the discipline of criminology cross-sectionally. That is, insulation may be attributable to the enterprise of discipline-building, which involves staking claims on the field's unique scholarly contributions. Thanks to Felice Levine for pointing this out.

7 Appreciation, though, can co-occur with marginalization, particularly when qualitative work is seen as an "added bonus" that, for instance, makes for attention-grabbing reading to attract the interest of distracted readers, but does not have epistemological impact on the core agendas of the discipline.

recently, also examining the impact of police actions on citizen behavior). In addition, official narrative records such as police incident reports have been employed in the study of homicide and other crimes. While each of these types of data sources would be useful for qualitative analysis of social processes associated with crime and criminal justice, they have been used primarily as source material to create quantitative datasets to examine event characteristics statistically.

Second is a general lack of training in, and thus understanding of, the epistemological underpinnings of qualitative research, appropriate qualitative data analysis strategies, and the particular strengths of qualitative research for illuminating social processes associated with crime and justice. Qualitative research is routinely characterized as "descriptive," "exploratory," and has even been excluded from what constitutes "empirical data." This has resulted in some seemingly contradictory outcomes.

On the one hand, because it is viewed as simply descriptive (and thus easy to accomplish), there is some use of qualitative methods among criminologists not trained in its methodology. The result is that qualitative data often appears as supplemental, descriptive data that provides "color" and "flash" to liven up quantitative analyses. Likewise, it is common to see studies that make use of qualitative data (most typically in-depth interview data) – but not qualitative analysis techniques – such that narrative accounts are fit within preconceived conceptual frameworks as descriptive evidence (or descriptive illustration), rather than being used inductively for theory building or theory refinement. Such work is problematic on a number of levels: it is unable to further social inquiry, because the data is not rigorously analyzed; and it reinforces a sense of the limited utility of qualitative inquiry for the development and elaboration of social theory. Often, analysis flaws (or lack of analysis) is apparent to a qualitatively trained eye, but goes unnoticed to a wider criminological audience.

On the other hand, the lack of understanding also means that qualitative studies are often held to inappropriate standards of (quantitative) rigor. For instance, this is the case when case studies, as well as purposive and snowball sampling strategies (typical and often necessary for the study of criminal offenders), are criticized for failing to adhere to standards of representativeness, and thus generalizability. In addition, misunderstanding of the method and its goals sometimes leads to a devaluation of its validity. For instance, I recently co-authored a paper based on in-depth interviews with African American young men about their negative experiences with the police. The paper was derided by one reviewer as "journalistic," and another reviewer referred to the data—in the case of young men's descriptions of their own encounters with the police—as "hearsay," and—in the case of their accounts of their perceptions of neighborhood policing—as "double hearsay." I give this example not to bemoan a set of negative reviews, but to highlight what I see as dangers to the qualitative enterprise when reviewers drawn from the broader discipline do not appreciate its methodological approach, and thus qualitative studies are expected to adhere to ill-fitting standards that undermine the very strengths and foundations of the methodology.

Standards of Rigor in Qualitative Research within Criminology

It is difficult to identify a uniform set of standards for rigor in qualitative research within criminology. This results in part from the issues noted above. On the other hand, I noted earlier a growing appreciation for qualitative research in the discipline. Despite the dominance of positivist paradigms adopting quantitative approaches, there are many significant qualitative studies of crime, criminalization, and criminal justice processes.

This research emanates from several sources, including scholars from other social science disciplines whose research addresses these issues but who do not participate actively in the discipline

(see, for example, Adler & Adler 1985; Bourgois 1995; Comfort 2003; Glassner & Loughlin 1987; Ferguson 2000; Katz 1988; Pattillo 1998; Venkatesh 2002), as well as criminologists (and criminological sociologists/anthropologists) themselves (see, for example, Anderson 1999; Fagan & Wilkinson 1998; Ferrell 1996; Fleisher 1998; Jacobs 1999; Maruna 2001; Moore 1991; Shover 1996; Sullivan 1989; Vigil 1988; Wright & Decker 1994, 1997). Qualitative research is perhaps most widely established in criminology among feminist scholars studying gender, race, crime and criminal justice[8] (see, for example, Bottcher 2001; Britton 2003; Joe-Laidler & Hunt 1997; Kruttschnitt, et al. 2000; Maher 1997; McCorkel 2003; Miller 1998, 2001; Miller & White 2003; Mullins & Wright 2003; Richie 1996). In addition, there have been several notable efforts among criminologists to integrate qualitative and quantitative research in ways that go beyond my earlier description[9] (see Hagan & McCarthy 1997; Nurse 2002; Sampson & Laub, 2003).

What does not exist, however, is ongoing, active dialogue among qualitative researchers in criminology to discuss issues of methodological rigor. Both critical mass and organizational infrastructure are largely absent. Nonetheless, I will make note here of some common themes and lauded works that provide insight into these issues. I focus specifically on qualitative studies in criminology that have utilized in-depth interview techniques, as this has been the approach most typically adopted (but see Ferrell 1996; Fleisher 1998; Maher 1997; McCorkel 2003). Much of this work focuses on what Hagan and McCarthy characterize as the "foreground" of crime, examining such issues as in situ motivations for offending, social processes associated with crime and the streets, and situational analyses of crime events. This is the primary body of work I will draw from here.

Sampling. Most qualitative research in criminology relies on purposive or snowball sampling techniques. The study of active offenders, for instance, requires the use of innovative methods to locate research participants. One technique successfully employed is the use of a fieldworker immersed in the social setting to generate contacts (see Jacobs 1999; Wright & Decker 1994, 1997). Another is the use of samples identified through various social control agents (Maruna 2001; Miller 2001), though this strategy comes with limitations that scholars must address in their analyses (see Agar 1977).

The Use of Comparative Samples. An increasingly popular sampling strategy within qualitative criminological studies is the use of comparative samples. Such an approach allows for some specification of similarities and variations in social processes and meaning systems across groups (for instance, offenders/non-offenders, desisters/persisters, females/males, African Americans/whites/Latinos), settings (neighborhoods, cities, institutional settings) and/or over time. This strategy can strengthen internal validity by allowing for more refined analysis and greater contextual specification.

Analysis Strategies.. This is perhaps the black box of qualitative research in criminology. While sampling and data collection procedures are widely discussed, systematic descriptions of the process of data analysis are typically not provided. Hagan and McCarthy (1995), for example, who blend qualitative and quantitative data, focus on the research questions of interest as their description of the analysis strategy. Some researchers describe the use of triangulation procedures, the examination of deviant cases, the use of interrater checks for reliability of analysis, and/or the use of tabular data to verify the representativeness of the patterns presented in data analysis. Such strategies, when described, speak to the internal validity of the analysis. Issues of reliability are typically addressed through the use of multiple interviews or repeated question sequences within a single interview, or are assumed by the current involvement of research participants in the activities of interest.

Finally, Maruna (2001) provides an example of a detailed description of his analysis plan,

8 Though within a discipline dominated by positivist epistemology, this results in some level of dual marginalization based on the combination of theoretical and methodological approaches.

9 It is also the case that an increasing number of leading quantitative scholars in the discipline have a favorable view of the import of qualitative research. While this is a cause for optimism, it has yet to result in a broader impact.

strongly influenced by positivist models. His discussion focuses on the use of blind coding by multiple coders (to achieve interrater reliability) of "episodes or phrases that were extracted from the body of the larger text" (p. 170) so that the coders had no information about the broader context of the interview. These pieces of text were then applied to "well validated" (p. 169) a priori coding schemes. Coding focused on "manifest (rather than latent) content or style" (p. 169), and a small number of women and minorities were included in the sample "in an effort to uncover the universal rather than the…specific" (p. 176).

I highlight this particular study in detail because Maruna's use of this analysis strategy was well received by quantitative scholars in criminology.[10] This raises vexing questions about the disciplinary push to inscribe such analytic strategies to a methodological approach whose strengths include inductive theory development, attention to latent meanings, and detailed attention to context, including the context of speech within the interview process (see Miller and Glassner, 2004). Because of the lack of transparency in most qualitative scholars' descriptions of their analysis techniques (and descriptions of the epistemological underpinnings of these techniques), there has not been sufficient critical debate to challenge the impetus to apply standards such as Maruna's systematically. This is exacerbated, as well, by the limited interchange more generally about qualitative work within the discipline.

Addressing Areas of Divergence Between Criminology and Other Disciplines

As described, standards of rigor in criminology are neither unified nor well specified. While this is less the case with regard to sampling and data collection techniques, it is a serious limitation with regard to standards for data analysis and the integration of theory. Several suggestions follow from this recognition.

First, it may be beneficial to develop avenues through which qualitative criminologists can come together to discuss methodological issues and address standards of rigor. This is important both to make the process more transparent for scholarly audiences, and to ensure that appropriate standards are applied in the evaluation of qualitative scholarship within the discipline. The lack of transparency is also associated with the limited amount of data sharing that takes place, and the general lack of public availability of the data collected for nearly all qualitative studies in the field.[11] In addition, I noted earlier the broader problem of criminology's insulation from theoretical and other developments in other social science fields. Greater cross-fertilization across disciplines would be beneficial for the field generally, and would benefit qualitative researchers specifically.

Second, a serious limitation within the field is the lack of qualitative training for young scholars. With few exceptions, the training of new generations of scholars within leading criminology programs is dominated by a quantitative theory-testing orientation. My own program, widely recognized as more diverse and complementary methodologically than most (with qualitative research methods a required course for the Ph.D. program), nonetheless strenuously trains students throughout the program in positivist epistemologies that emphasize quantitative theory modeling and testing. Of course, these are practical imperatives of the market. But the broader problems for qualitative scholarship within the discipline are nonetheless replicated. Anecdotally, I would suggest this does not stem from a lack of interest on students' part. I am routinely invited to speak with students in programs in which there is recognition among faculty of student interest in the methodology, but limited faculty specialization to meet this need.

---

10 His study received the major book award in 2001.

11 Of course, this is a challenging problem for criminology in particular because the protection of confidentiality is especially important given the potentially damaging nature of data collected from those involved in crime.

Promising Areas for Investigation in Criminology

Several topics and areas are promising in the discipline for investigation. Qualitative methods have been applied widely in situational studies of crime, and have included analyses of such issues as offender decision-making, social networks, and social processes shaping criminal events. In some cases, this research has been combined with analyses of gender/race inequalities, as well as neighborhood and street contexts. Such studies will be improved with greater attention to strategies for making linkages between micro-level interactional processes, state policies, and structural inequalities (see Wacquant 2002). Again, this would be facilitated by decreased insulation from developments in other social science disciplines.

Additional areas have received some limited attention from qualitative researchers, but offer promise for future investigations. This includes the use of life history narratives to understand pathways to offending (and desistance) (see Daly 1992; Gaarder & Belknap 2002; Giordano et al. 2002; Moore 1991; Sampson & Laub 2003), as well as research on organizational processes and decision-making within criminal justice and other relevant institutions (see Britton 2003; Comfort 2003; Ferguson 2000; Frohmann 1991; Kruttschnitt et al., 2000; McCorkel 2003).

In addition, a number of opportunities exist for greater cross-fertilization within the discipline between qualitative and quantitative research. For example, there are promising examples of efforts to test theoretical propositions emerging from qualitative studies using quantitative analyses (see Peterson et al. 2001; Stewart & Simons 2004). Moreover, efforts are underway to develop studies that are multi-method, multi-level and/or comparative, and these offer promising approaches for efforts to integrate qualitative and quantitative research (see for example Hagan & McCarthy 1997; Klein et al. 2001). Finally, there are underdeveloped opportunities with data already on hand for such combined efforts: quantitative studies which draw from qualitative data, such as those I made note of earlier, offer just such an untapped resource.

References

Adler, Patricia A. and Peter Adler. (1985). *Wheeling and Dealing.* New York: Columbia University Press.

Adler, Patricia A. and Peter Adler. (1987). *Membership Roles in Field Research.* Newbury Park: Sage Publications.

Agar, Michael H. (1977). "Ethnography in the Streets and in the Joint: A Comparison." Pp. 143-156 in *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings,* edited by Robert S. Weppner. Beverly Hills, CA: Sage Publications.

Anderson, Elijah. (1999). *Code of the Street: Decency, Violence and the Moral Life of the Inner City.* New York: W.W. Norton.

Bottcher, Jean. (2001). "Social Practices of Gender: How Gender Relates to Delinquency in the Everyday Lives of High-Risk Youths." *Criminology,* 39: 893-932.

Bourgois, Philippe. (1995). *In Search of Respect: Selling Crack in El Barrio.* Cambridge: Cambridge University Press.

Britton, Dana M. (2003). *At Work in the Iron Cage: The Prison as Gendered Organization.* New York: NYU Press.

Comfort, Megan L. (2003). "In the Tube at San Quentin: The 'Secondary Prisonization' of Women Visiting Inmates." *Journal of Contemporary Ethnography,* 32: 77-107.

Daly, Kathleen. (1992). "Women's Pathways to Felony Court: Feminist Theories of Lawbreaking and Problems of Representation." *Review of Law and Women's Studies,* 2: 11-52.

Fagan, Jeffrey and Deanna L. Wilkinson. (1998). "Guns, Violence and Social Identity in Inner Cities." Pp. 105-188 in *Youth Violence,* edited by Michael Tonry and Mark H. Moore. *Crime and Justice,* 24. Chicago: University of Chicago Press.

Ferguson, Ann A. (2000). *Bad Boys: Public Schools in the Making of Black Masculinity.* Ann Arbor: University of Michigan Press.

Ferrell, Jeff. (1996). *Crimes of Style.* Boston: Northeastern University Press.

Fleisher, Mark S. (1998). *Dead End Kids: Gang Girls and the Boys They Know.* Madison: Wisconsin University Press.

Frohmann, Lisa. (1991). "Discrediting Victims' Allegations of Sexual Assault: Prosecutorial Accounts of Case Rejections." *Social Problems,* 38: 2.

Gaarder, Emily and Joanne Belknap. (2002). "Tenuous Borders: Girls Transferred to Adult Court." *Criminology,* 40: 481-518.

Giordano, Peggy, Stephen A. Cherkovich and Jennifer Rudolf. (2002). "Gender, Crime and Desistance: Toward a Theory of Cognitive Transformation." *American Journal of Sociology,* 107: 990-1064.

Glassner, Barry and Julia Loughlin. (1987). *Drugs in Adolescent Worlds: Burnouts to Straights.* New York: St. Martin's Press.

Hagan, John and Bill McCarthy. (1997). *Mean Streets: Youth Crime and Homelessness.* Cambridge: Cambridge University Press.

Joe-Laidler, Karen A. and Geoffrey Hunt. (1997). "Violence and Social Organization in Female Gangs." *Social Justice,* 24: 148-169.

Jacobs, Bruce A. (1999). *Dealing Crack: The Social World of Streetcorner Selling.* Boston: Northeastern University Press.

Katz, Jack. (1988). *Seductions of Crime: Moral and Sensual Attractions of Doing Evil.* New York: Basic Books.

Klein, Malcolm W., Hans-Jurgen Kerner, Cheryl L. Maxson and Elmar E.M. Weitekamp (eds.) (2001). *The Eurogang Paradox: Youth Groups and Gangs in Europe and America.* The Hague: Kluwer

Kruttschnitt, Candace, Rosemary Gartner and Amy Miller. (2000). "Doing Her Own Time? Women's Responses to Prison in the Context of the Old and the New Penology." *Criminology,* 38: 681-718.

Maher, Lisa. (1997). *Sexed Work: Gender, Race and Resistance in a Brooklyn Drug Market.* Oxford: Clarendon Press.

Maruna, Shadd. (2001). *Making Good.* Washington D.C.: American Psychological Association.

McCorkel, Jill A. (2003). "Embodied Surveillance and the Gendering of Punishment." *Journal of Contemporary Ethnography,* 32: 41-76.

Miller, Jody. (2001). *One of the Guys: Girls, Gangs and Gender.* New York: Oxford University Press.

Miller, Jody. (1998). "Up It Up: Gender and the Accomplishment of Street Robbery." *Criminology,* 36: 37-66.

Miller, Jody and Barry Glassner. (2004). "The 'Inside' and the 'Outside': Finding Realities in Interviews." Pp. 125-139 in *Qualitative Research,* 2nd edition, edited by David Silverman. London: Sage Publications.

Miller, Jody and Norman A. White. (2003). "Gender and Adolescent Relationship Violence: A Contextual Examination." *Criminology,* 41: 1501-1541.

Moore, Joan. (1991). *Going Down to the Barrio: Homeboys and Homegirls in Change.* Philadelphia: Temple University Press.

Mullins, Christopher W. and Richard Wright. (2003). "Gender, Social Networks, and Residential Burglary." *Criminology,* 41: 813-840.

Nurse, Anne. (2002). *Fatherhood Arrested.* Nashville: Vanderbilt University Press.

Pattillo, Mary E. (1998). "Sweet Mothers and Gangbangers: Managing Crime in a Black Middle-Class Neighborhood." *Social Forces,* 76: 747-774.

Peterson, Dana, Jody Miller and Finn-Aagee Esbensen. (2001). "The Impact of Sex Composition on Gangs and Gang Member Delinquency." *Criminology,* 39: 411-439.

Richie, Beth E. (1996). *Compelled to Crime: The Gender Entrapment of Battered Black Women.* New York: Routledge.

Sampson, Robert J. and John H. Laub. (2003). *Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70.* Cambridge, MA: Harvard University Press.

Shover, Neal. (1996). *Great Pretenders: Pursuits and Careers of Persistent Thieves.* Boulder, CO: Westview.

Stewart, Eric A. and Ronald L. Simons. (2004). "Structure and Culture in African American Adolescent Violence: A Partial Test of the Code of the Street Thesis." *Ms.*

Sullivan, Mercer L. (1989).*"Getting Paid": Youth Crime and Work in the Inner City.* Ithaca, NY: Cornell University Press.

Venkatesh, Sudhir Alladi. (2002). *American Project.* Cambridge, MA: Harvard University Press.

Vigil, James Diego. (1988). *Barrio Gangs: Street Life and Identity in Southern California.* Austin: University of Texas Press.

Wacquant, Loïc. (2002). "Scrutinizing the Street: Poverty, Morality and the Pitfalls of Urban Ethnography." *American Journal of Sociology,* 107: 1468-1532.

Wright, Richard and Scott Decker. (1994). *Burglars on the Job: Streetlife and Residential Break-ins.* Boston: Northeastern University Press.

Wright, Richard and Scott Decker. (1997). *Armed Robbers in Action: Stick Ups and Street Culture.* Boston: Northeastern University Press.

In Search of Social Science

Susan Silbey
Department of Anthropology
Department of Sociology
Massachusetts Institute of Technology

"You might be sitting in a history class/ listening to the analysis of 'what was going on" in the thirties in new york, say/ and you hear nothing of shtetls where grandma's generation came from/ and the descriptions of sweatshops sounds oddly abstract and disembodied, that is, emphatically unsweaty-scientific-full-of clear-light - spared of the dust of ripped cloth - and quiet so you can hear yourself (or someone else) think and the machines screaming and bobbing has stopped, all put in terms of an analysis of the labor structure of the immigrant population, complete with statistics/ and politics sounds like this or that labor policy adopted by this or that administration/ not at all what grandma described doing to work as/ but you came to school to learn/ and it feels like an interesting addition to what you already know from family history and hot tea mornings in kitchens in brooklyn apartments/ but it still seems like the viewpoint of the other, of the officials giving the official line on what was happening - the politics at the pinnocle games just can't be reduced to "labor unrest"/ but we're going too fast"

"then its years later and you wonder again about the shtetls and what you might have lost in the history class/ and you focus on some imaginary moment when it happened - when the statistics and the analysis of the labor structure were no longer just interesting additions to the lived experience in new york of grandma and her friends but instead became the reality itself; and grandma's description about why her boss acted like he did was just shtetl superstition, or worst, silly, because at some point the feeling of learning new things was replaced by the idea of learning things they way they really are, free from superstition and prejudice, and stuff might be left out for the sake of time but what was there, presented as knowledge, was knowledge, in a particular form and in a particular language that you recognize as not the way you started out looking at things, but we're for education after all."

"and then you start wondering what if the language of true knowledge that you learned, the way of talking about things intelligently and dispassionately, was itself a mythology that contained prejudice and superstition; and then that its not just new york in the thirties, its the way the whole picture is organized, a whole hierarchy of what counts and what doesn't count that might present itself as neutral knowledge but is really just an ideology of power/ and the imaginary moment that you crystallized, the moment when the statistics and the analysis began to represent the true and the real against the superstitious, was the moment of self-denial and treachery as you implicitly agreed to a test of truth that would count out most of what you knew most deeply, even if you can't prove it."

-Gary Peller, "Violence and the Mob" Tikkuyn, 1987

This self-conscious and stylized reconstruction of a supposedly spontaneously produced, authentic memory is the introduction to Gary Peller's 1987 essay, "Reason and the Mob: The Politics of Representation." In this essay, appearing in Tikkun magazine in its first year, Peller laid out the claims for, and methods of, deconstruction as a mode of interpretation and critique. Nineteen eighty seven was a heady moment in legal scholarship, at least for the community of progressive legal scholars. Critical Legal Studies - of which Peller was a major player - was at its zenith and moving out into the popular culture, with articles in such venues as the Nation, the New Yorker, New York Times, and the New Republic. The media attention - whether positive or critical - was a sign that this radical critique of law and legal institutions had reached beyond the ivory tower, out of the vaulted law school corridors.

At the same time, the cultural turn was gaining momentum, gathering adherents, developing scholarly potential, as well as securing occupational niches. Interdisciplinarity was emerging as the new buzzword, not only among scholars working at the boundaries of their fields but even among administrators and disciplinary strongholds.

Similarly, ethnographic and qualitative analyses were expanding out of anthropology and the backwaters of sociology into quite diverse fields; and, importantly, from my perspective, attention to the meaning of action, not only its distribution, was becoming more and more important as an antidote to mindless quantification and ideological modeling in the heart of what claimed to be value free social science.

Many legal scholars, and social scientists, were particularly impressed by the erudite performances of literary scholars, for their ability to unearth complex and subtle meanings in representations of human action. And thus, arguments about the connections between literature and social science, literature and law, literature as a map of social life were becoming increasingly persuasive. Literary criticism seemed, to some, to offer a way of both accessing and representing the new interest in culture. Some describe this intellectual moment as a significant rupture in modernity and modern scholarship, others refer to it in terms of warfare: the culture and science wars.

At that moment, Peller's didactic essay offered - at least to me - an entree, a window on what seemed like important transformations, and powerful insights. I chose Peller's little stream of conscious representation to introduce my comments for this workshop because when I first read it years ago, it seemed emblematic to me of my own journey as a scholar, why I teach research methods, and why I happily study in the field of law and society. I was, in some ways like Peller, someone who began from very similar circumstances - also listening to grandma and grandpa in the kitchen, captivated by their energy and passion, warmed by the mist of tea served in tall glasses, and intrigued by the stories of political battles I could not really understand. Peller's account seemed familiar to me in terms of the family dynamics he suggested, and indeed in terms of the seduction and transformation in consciousness and capacity that his college education seemed to produce - at least until his subsequent rejection of it. And like Peller, I too was attracted to literary theory and to what seemed like powerful methods for tracking the complex meanings in human relations.

But, despite significant parallels in Peller's account and one I might construct of my own childhood and education, there are also big differences, differences that derive, I believe, from the attachment, confidence, and pleasure I find in social science, especially a social science of law, and in the kind of public scholarship and engagement it signifies for me. I believe that social science offers the possibilities of just that engagement, connection, reason, and kind of truth telling that Peller sought, but somehow failed to find.

Unlike Peller, I did not experience my college education – way back then when I was in college or now as I reflect upon it, nor my graduate education for that matter – as alienating abstraction, disembodied, unsweaty, or as self-denial or treachery of things that I knew most deeply. Nor have I come to understand what I learned in college as just other mythologies, just more prejudice and superstition. Neither, by the way, did my education encourage me to reject connections to my family's world.

Quite the opposite. I experienced my college education, and much of my work ever since, as a comforting liberation. I say liberation not only because I was freed from unwanted constraints, although I certainly was. Nor was it libratory because I perceived something "full of clear light," as Peller says. Nothing was clear - not my grandfather's politics, nor my professor's lectures. I believe that my education was both liberating and comforting because I was offered an alternative world, a new kind of security, something that happily substituted for the warm tea in tall glasses and family chatter in the kitchen, but which also happily lacked the unreasoned authority, prejudice, sentimentality and superstitions that confounded and oppressed me as a young girl of moderate intelligence, insufficient beauty, and too much energy growing up in Brooklyn NY in the 1950s and 60s.

By my young middle age, I was grateful for not having been an adolescent star within my family or school; I came to believe and still do, that my outsider status as a child and young woman made me a much more astute observer of social situations and structure. Unlike the confining world of my family, what I learned in college was more general than idiosyncratic, more collective than secret, frankly less lonely, more predictable, and stable than what passed for received wisdom in my household. This new world validated me, my experiences of fairness and unfairness, my intuitions about the injustices of powerful authorities, the banality and oppressiveness of gendered expectations.

I grew up in a world where salt thrown over your shoulder protected you from the evil eye (I was never sure whose evil eye), where those evil eyes could cause financial as well as bodily damage, where wet bathing suits caused kidney problems, where young girls who were not married at twenty were problems for their family, where women did not have interests in politics, or justice, or what we would call the public sphere, and where any woman who had ideas (and voiced them) about politics and the public sphere was simply not a good woman. I am not making a specifically feminist point here. I am merely using my experience to illustrate what I know many men also experienced - living in worlds they could not understand nor influence very well, but which they both loved and felt outside of.

Along with the warm fuzzy memories, I had also experienced my childhood, as I have said, as tyrannical, cruel and unfair. A large part of it was unfair because, I did not seem to have any influence, or what we might call voice, and because it seemed unpredictable. One could never really know why one thing was good or bad, preferable or to be avoided. Why could I drive my car - and I was old enough to drive and privileged enough to have my own car - could drive it anywhere within the boundaries of NYC but never cross the Hudson River; that restriction meant I could go to Harlem, Brownsville, the highest crime neighborhoods within the city, but not to the Princeton University bookstore that had just the book I wanted. These rules seemed like and were expressions of the blatant, absolute power of merely older persons whose authority was constituted solely because they were older - parents, teachers - none of whom seemed to see who I was. I was one of those obnoxious, but angry little children who would regularly decry to these powerful figures, including my older sister, “just because you are older, doesn’t mean you are smarter or know more.” I was their little problem, and they and I knew it.

So, when I got to college and discovered political science and then sociology and then anthropology, I thought I had found nirvana. Here, it not only was possible but was expected to have arguments about what was just or fair, who should be heard and how they should be heard, what was true, what was not plausible, and what constituted the good. I loved studying philosophy and when I discovered the logical positivists and analytical philosophers, I thought I had been handed Excalibur. I thought that through empirical research, we could come to shared conversations, if not conclusions on these questions.

Unlike Peller, my college education transformed me but did not alienate me. It offered a new home, one that was less partial, less particular, less idiosyncratic than the world of my family. And because this new world of social science was more open and becoming increasingly transparent to me, and because uncertainty was recognized and valued, it was libratory. lt was libratory because it offered the security of community, of boundaries, and yet of action and imagination - what I think of as method. Not anything I said or wrote, or anyone else, would constitute a reasonable account. There were shared expectations of how to go about constructing one’s accounts of the world and those methods - quite diverse and varied, including surveys and models, and narratives, and deep ethnographies - were open to inspection and interrogation. They were shared, collective and public, debated, subject to imaginative reconstructions and deployments.

My point is simply this: I believe that knowledge construction occurs collectively, and

publicly. Indeed, I do not simply believe this but I think we can show this. Thus, the collective, public constitution of science, including social science, has a built in politics, a built-in justice. What we know, through normal science, we know because we have created it through a relatively open, and I stress relatively not absolutely, participatory community. However, to the extent that we - as scholars - do not explain to our audience how we produced our accounts, we might as well return to the world of authority, or unreason. To the extent that we hide the ball so to speak, we are engaged in a more rather than less, private and personal activity. We are left to perform or display ourselves, a solipsistic, perhaps even narcissistic self-referential mirror, but we do not offer knowledge to be shared, exchanged, acted upon. To the extent that we study law, with its aspirations for justice, I think we have an obligation to enact those very same collective aspirations through our methods of study - open, transparent, participatory, thus public and accessible.

Liberation consists of knowing the rules, whether of a powerful apparatus such as the law, that is rules of public coercion, or the rules of reason and knowledge production. I imagine that Peller might have been disappointed in the enlightenment of his education because it did not reveal the rules of its construction. It might as well have been revealed truth. Thus, I was drawn to Peller's account because we shared an attraction for literary techniques and Peller, unlike many, does not simply perform his deconstructive reading, but he instructs his readers on how to do it, he does not hide the ball. He does not offer revelation, but demonstration. Thus he taught me and I was grateful. But, unlike Peller, the more I encountered and was surrounded by interpretive scholarship that was, for want of a better word, performative, the more I longed for normal social science. By normal social science, I do not mean mindless counting, or atheoretical description. By performative scholarship, I refer to what are often engaging, artful, seductive displays of erudition, insight, and playfulness that are offered without a display of or map of their construction.

Thus, my thought about how to improve qualitative methods is simple and banal: make them transparent, as much as we can. Of course, there will always be leaps of intuition, serendipity, inferences that are not entirely supported, but we should, nonetheless, explain to our readers, and especially to our students, how we came to the conclusions we offer. In this regard, I have several worries and suggestions:

(1) Teach more qualitative methods courses. I recently surveyed all the major sociology and anthropology departments to determine whether they offered courses in qualitative methods and/or ethnography. It turned out that although several universities across the nation provide excellent training in fieldwork and ethnography, it is not as common or consistent as is the preparation in quantitative data collection and statistical analyses. The major sociology departments offered at least one but did not require such a course although they all required courses in general research methods, another in statistics and a range of offerings in advanced quantitative analyses. Anthropology departments did not regularly offer nor require fieldwork methods, and political science rarely offered courses in qualitative data collection or analysis. Although case studies and historical studies are not unusual, the methods courses were more often survey methods. So my first worry is that we cannot create more transparent methods if we do not teach them as regularly, or consistently.

(2) Pay attention to reliability and validity. My second worry is that in the well-founded critique of positivistic behaviorism, qualitative scholars have too easily ignored collective and reasonable criteria for reliability and validity. In rejecting a correspondence theory of truth, we have left ourselves open to criticism and rebuke. Thus, although the use of qualitative research methods has grown in recent decades and writing about qualitative methods has also proliferated, there has been "no parallel proliferation of studies of the actual process of doing qualitative research (Huberman and Miles 2002:x). The

cause of this inattention to the process is over-determined, a product of ethnography's own history, the epistemological debates between different approaches within qualitative methods, no less between qualitative and quantitative researchers, as well as the culture and science 'wars' of the post-structural turn in the social sciences I mentioned at the outset. As a consequence, the processes of data collection and analysis, as well as the distinctions and connections between these, are not well understood, especially among non-practitioners. The publication of many new texts and handbooks does not seem to have helped. Often, one observes a general lack of credibility for the results of qualitative research. When researchers "spend more pages explaining why they will not deploy particular methods than on describing their own conceptual and analytic moves, they leave behind too few footprints to allow others to judge the utility of the work, and to profit from it" (Huberman and Miles 2002:xi). Thus, it seems important for improving qualitative methods that we begin to unpack the process and make it more transparent. This is as important for the collection of data as for the analysis, which in ethnographic fieldwork is always continuous and simultaneous with data collection (Becker 1998, 2004; Silbey 2004).

One of the most frequently voiced concerns about ethnography is whether a different observer would have come away from the field, independent of the variations in the voice with which the account may be written, with the same basic account. In other words, how reliable is this description of the social world depicted in ethnographies. In ethnographic research, however, reliability is closely connected and perhaps best understood as a form of validity (Hammersley 1992:79). Although these terms are normally reserved for quantitative and positivistic research, I use them to refer to the ability to produce similar data from multiple observers, and to produce consensually agreed upon, corroborated, accounts of a site, person, or process. They can be deployed for ethnographic research with some modifications. In a succinct account of these, Maxwell (1992) for example, proposes five types of validity for qualitative researchers that offers an advance on the usual discussions of reliability and validity. *Descriptive validity* refers the factual accuracy of an account, that researchers "are not making up or distorting things they saw or heard" (1992:45). This is the basis for all other forms of validity. As Geertz (1973:17) put it, "behavior must be attended to, and with some exactness, because it is through the flow of behavior – or more precisely, social action – that cultural forms find articulation." This "reportage" (Runciman 1983) refers to specific events and situations, as well as to objects and spaces. As Maxwell (1992:46) says, " no issue of generalizability or representativeness is involved. These are matters on which, in principle, intersubjective agreement could easily be achieved, given appropriate data."

Interpretive validity refers to representations of what the described behaviors, events, and objects mean to the actors observed, one of the central goals of qualitative research, especially ethnographic fieldwork. Interpretive validity seeks, in one conventional framing of qualitative research, to capture the participants' perspective, description in emic terms. Interpretive validity "has no real counterpart in quantitative-experimental validity typologies… [It] is inherently a matter of inference from the words and actions of participants in the situations studied… grounded in the language of the people studied, [and] relying as much as possible on their own words and concepts" (Maxwell 1992:48). The goal of interpretation is to describe the actors' "lay sociology" (Garfinkel 1964) or "theories-in-use" (Argyris and Schoen 1978). This criterion of interpretive validity distinguishes a form of accuracy that lies between the first form, descriptive validity, resting entirely on observable data and the more contestable interpretations of the third, theoretical validity, to which I will turn next. While there is "no in-principle access to data that would unequivocally address threats to [interpretive] validity," the descriptive accounts serve as warrants, and consensus should be achievable within the relevant community of actors about "how to apply the concepts and terms of the account" (Maxwell 1992:49). The concepts and terms of both descriptive and interpretive validity are, in Geertz's term, "experience-near,' the local

language in use among the actors, although interpretive validity might also involve assessments of the accuracy of informants' reports. "Accounts of participants' meanings are never a matter of direct access, but are always constructed by the researcher(s) on the basis of the participants' accounts and other evidence" (Maxwell 1992:49).

Theoretical validity moves the ethnographic account further from the actors' behavior, language, meanings, and interpretations to a more abstract account that proposes to explain what has been observed. "Theoretical validity thus refers to an account's validity as a theory of some phenomenon" (Maxwell 1992:51). Both the concepts used and the relationships proposed are independently assessed for what is conventionally called construct validity (Bernard 2000:50-51) and inferential or causal validity (Cook and Campbell 1979), although not all theories proposal to explain causes. The key distinction between these types of validity, in this schema, lies in the "presence or absence of agreement within the community of inquirers about the descriptive or interpretive terms used. Any challenge to the meaning of the terms, or appropriateness of their application to a given phenomenon, shifts the validity issues from descriptive or interpretive to theoretical" (Maxwell 1992:52). *Generalizability* is refers to "the extent to which the particular situation is representative of a wider population" (Hammersley 1992:79). There is, however, a level of analysis issue here concerning generalizability that distinguishes internal from external validity. To what extent were the observed persons and activities representative of that particular group or organization? This "internal generalizablity is far more important for most qualitative researchers than is external generalizability because qualitative researchers rarely make explicit claims about the external generalizability of their accounts." As Friedson (1975) writes in warranting qualitative research, "there is more to truth or validity than statistical representativeness." Nonetheless, we need to be careful about the claims made, implicitly and explicitly, for generalizability, internal and external. Qualitative researchers normally elide external generalizabilty by offering the particular case as an example from which to generate typologies and hypotheses rather than test theories. Maxwell offers a final fifth form of *evaluative validity*, referring here to the normative assessment of that which has been described or explained. This category is perhaps most appropriate for policy studies and are not intrinsically different in qualitative than quantitative studies.

(3) Make social science a collective, public enterprise. Teach and do fieldwork in groups. One of the special features of graduate education in science is its collective form. Scientists work in groups. Laboratory groups vary in size: number of actual rooms and researchers. The research done under each professor's supervision is conventionally referred to, depending on the particular speaker, as "a lab," "our lab" or "Johnson's group," invoking the name of the faculty supervisor. The faculty member or principal investigator bears responsibility for designing the research agenda, for selecting the researchers (technicians, post-doctoral fellows, graduate and undergraduate students). Laboratory groups normally meet once or twice a week to discuss work collectively, to present results to each other, to comment on papers. Group meetings are key for developing the interpretations of data, as well as the camaraderie in the group and the authority of the supervising faculty (Owen-Smith 2001). One's identity as a scientist is tied intimately to this group and its advisor, with successions of faculty and students describing themselves as kin, often publishing their genealogies on their web pages. Although the genealogies highlight, in bold color, a Nobel prize in the line, it is not unusual to find such lineage charts going back much further in history long before Nobel, securing a direct link between a contemporary scholar and one of the 17th century founders of modern science. The laboratory group is a family, not only as source of identity and authority but also as a supporting network. Social scientists enjoy this support less often, and humanists almost not at all. We would do well to consider the benefits that accrue to teaching students in groups, meeting and discussing research in groups. In this way, we would demonstrate in the very training of social scientists that this is a public, collective enterprise, not solely

a matter of individual talent, insight, or creativity. We would, by working and teaching in groups, create the conditions for articulating the grounds of how we know what we know.

References

Becker, H. (2004). *The Problems of Analysis*. Pp. 45-48 in *Workshop on Scientific Foundations of Qualitative Research*, edited by C. Ragin, J. Nagel and P. White. Arlington, VA: National Science Foundation.

Bernard, H. R. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. Beverly Hills: Sage Publications.

Cook, T. D., and D. T. Campbell. (1979). *Quasi-experimentation; Design and AnalysisIssues for Field Setting*s. Chicago: Rand McNally College Publishing.

Freidson, E. (2001). *Professionalism: The Third Logic*. Cambridge, UK: Polity.

Garfinkel, H. (1964). "Studies of the Routine Grounds of Everyday Activities." *Social Problems*, 11: 225-50.

Geertz, C. (1973). "Thick Description: Toward an Interpretive Theory of Culture." Pp. 3-30 in *The Interpretation of Cultures*, edited by C. Geertz. New York: Basic Books.

Hammersley, M. (1992). *What's Wrong with Ethnography*. London: Routlege.

Huberman, A. M., and Miles M. B., eds. (2002). *The Qualitative Researchers' Companion*. Beverly Hills: Sage Publications.

Maxwell, J. A. (2002). "Understanding and Validity in Qualitative Research." Pp. 37-64 in *The Qualitative Researcher's Companion*, edited by A. M. Huberman and M.B. Miles. Beverly Hills: Sage Publications.

Owen-Smith, J. (2001). "Managing Laboratory Work Through Skepticism: Processes of Evaluation and Control." *American Sociological Review,* 66: 427-452.

Peller, G. (1987). "Violence and the Mob." *Tikkun*.

Silbey, S. S. (2004a). Designing Qualitative Research Projects. Pp. 121-126 in *Workshop on Scientific Foundations of Qualitative Research*, edited by C. Ragin, J. Nagel and P. White. Arlington, VA: National Science Foundation.

Appendix 5.

Political Science

Defining Excellence in Qualitative Research: Groups Report for Political Science

A Note for Evaluators: Differences from Quantitative Research

Qualitative methods are useful at all stages of the research cycle, including both theory development and theory testing. Judging the use of qualitative methods requires sensitivity to the justifications of and standards for using these methods, which differ from those of standard frequentist statistical methods on several key issues.

Case Selection: Allowing for Causal and Relational Complexity

Assessing qualitative work requires being attuned to a wide variety of kinds of possible causal relationships. Causation may involve probabilistic relations and linear or other simple functional forms, which are amenable to linear regression analysis, but casual relationships may also include forms of complexity that are often easier to assess through qualitative analysis and comparison of one or a few cases. A single case study, for example, can cast doubt on claims that a condition is necessary or sufficient for an outcome. Also, when multiple paths are hypothesized to the same outcome, a form of causal heterogeneity known as "equifinality," case studies can map out these different paths. Qualitative case studies are also particularly effective at assessing complex interaction effects, as the detailed evidence available within particular cases allows analysis of many possible interactions, and path dependencies, which qualitative researches can analyze through process tracing on the sequences of events within cases. The qualitative methods that allow inferences about these kinds of complexity include in-depth interviewing, process-tracing, typological theorizing, ethnography, textual analysis, and archival research. These methods involve different modes of inference from those commonly promulgated in textbooks on statistical methods and do not always rely on frequentist logic; a single interview or text, for example, may show definitively that a hypothesized process did not take place in a case and cannot explain the case, even if a large number of other observations are consistent with the hypothesized explanation of the case.

The Value of 'Thick Description'

Moreover, while many qualitative researchers aspire in much of their work to the goals of theoretical generalization and prediction, qualitative researchers also value precise descriptions. Accurate descriptions and measures are essential to both qualitative and quantitative means of inference. Innovative descriptions, which show that a case or instance does not fit a concept or category to which it is typically assigned, are important as well. Such descriptions are typically based on highly detailed knowledge of one or a few cases, and they contribute directly to new concepts. Such descriptions may be carried out either in relation to frames provided by the investigator and/or by the actors being studied, the latter involving interpretation. Qualitative researchers also value the use of theories to develop historical explanations of particular cases. Research focused on historical cases involving a distinctive constellation of factors – for example the Cuban Missile Crisis, the rise of Nazism – can help to identify the more general conditions under which crises are successfully de-escalated or fascist parties rise to power. In addition to using theories to understand such cases, qualitative researchers can use historical cases to further develop and test theories.

Standards for Evaluators and Researchers

Keeping in mind that the goals and means of qualitative research are diverse, and that the standards for judging any one research project cannot be divorced from consideration of its goals, its relationship to the extant literature, and the nature and availability of the evidence on the phenomenon under study, excellent qualitative research usually embodies the following standards:

Project Framing

- **Relevance to real-world problems.** Good qualitative research may be useful in diagnosing and addressing real-world problems, and if intended for that purpose, can readily be made intelligible to the actors and policy-makers involved in addressing such problems.
- **Intellectual relevance.** Counter-intuitive findings that call into question established verities are an important goal of research. Excellent research should be well-situated in relation to the literature on a topic, but often it also stands in some tension to the accepted wisdom.

Research Design and Methodology

- **Careful attention to case selection.** Cases should be appropriate for the research objective, the kinds of comparisons to be drawn, and the type or population to which generalizations might be extended. In addition to specifying which cases they chose to study and why, researchers should indicate which cases they considered for inclusion but ultimately decided not to study in depth, and explain why they set these cases aside. Researchers should also indicate to which populations or types of cases their findings should or should not apply.
- **Validity.** Internal validity, in which the measures used to represent the specified concepts do indeed faithfully represent these concepts, and external validity, in which the arguments the researcher develops or tests are applicable to the population they specify, are both important to qualitative research. Some individual research projects (for example those that focus on either intensive description or historical explanation of one or a few cases) may downplay the goal of applying their findings to wider populations, and others may propose generalizing only to narrow and well-specified sub-types of a phenomenon. In some research projects, investigators may seek to find or happen to uncover evidence on causal mechanisms relevant to wide and diverse populations. The important criterion here is that good qualitative research should be clear on the mix of historical explanation and theoretical generalization to which it aspires, and provide evidence of having attained it.
- **Replicability**. The description of the methodology should be sufficiently detailed and transparent to allow other researchers to check the results and/or reanalyze the evidence. All reasonable efforts should be undertaken to make archival and other data available to scholars.

Data Analysis

- **Attention to the possibility of reciprocal causation and identification problems.** Generally, researchers should explore whether the causal factors of interest are spurious substitutes for the operation of other, un-measured causes. At the same time, reciprocal relationships (feedback effects or expectations effects, for example) may themselves be the focus of qualitative research.
- **Iteration between theory and data** is often a central component of qualitative research. The most fruitful and defensible iterations involve the generation of novel insights or facts

that are then tested against additional independent evidence from a case or from other cases. This helps reduce the risks of spurious inferences and confirmation biases.

- **Falsifiability, broadly construed.** While a narrow, simple, or "naïve" standard of falsifiability is not advisable and has been thoroughly critiqued in the philosophy of science, a broad view of the relevance of negative evidence is invaluable. This is important as a means of responding to potential objections and as a guard against confirmation biases.
- **Attention to alternative explanations.** Demonstrating that alternative explanations do not fully account for the outcome of interest, and/or that the main explanation of interest subsumes, complements, or supersedes the alternative explanations.

Presentation

- **Logical coherence**. The theories, descriptions, or explanations developed or tested through good qualitative research are logically coherent and internally consistent.

Promising Substantive Topics for Qualitative Research

In general, there are few substantive research programs within political science to which qualitative methods cannot contribute. These methods are likely to be especially useful, however, in the development and testing of theories about phenomena that are relatively new, infrequent, or complex, or that combine some or all of these three characteristics.

- In American politics, qualitative research might fruitfully focus on the rise of religious movements, relations between racial/ethnic and class identities and political action, and policy-making processes, especially those that cut across different institutions.
- In comparative politics, qualitative research on democracy and democratization could be expanded to move beyond minimalist definitions to include the ways in which everyday democratic practices operate within authoritarian regimes. Studies of democracy might also explore the various meanings of the concept in different cultural and historical contexts.
- In international relations, fruitful areas for qualitative research include the role of non-state actors, terrorism and other non-traditional forms of combat, financial crises, proliferation of WMD-weapons of mass destruction, the effects of global trade on local communities, and competition among and with emerging great powers.

Producing Top Notch Qualitative Research in Political Science: The NSF's Role

As the most significant funding organization for the social sciences, the NSF can have a substantial impact on enhancing the quality of qualitative research produced across disciplines. In particular, the NSF might usefully take the following steps to further promote qualitative research in political science and related disciplines:

Training Programs in Qualitative Research

One primary means the NSF can encourage excellence in qualitative research is by providing scholars with enhanced training in qualitative methodology. Such training programs should be available to scholars at all career stages, including graduate students, and encompass both the practice and teaching of qualitative methods. Examples of such initiatives include:

- Continuing to fund the annual Arizona State Institute on Qualitative Research Methods (IQRM).

This institute has involved several dozen faculty over its first four years and is currently training approximately eighty advanced graduate students and junior faculty in qualitative research methods over the course of two weeks each year. NSF funding, which is about 20% of the institute's budget, allows the institute to reach beyond the member departments and research centers that comprise most of the institute's budget and to include students from a diverse range of non-member institutions.

- Funding a qualitative methods component or follow-up project to EITM (Empirical Implications of Theoretical Models) exploring the relationship between formal theory and qualitative methods
- Sponsoring the development of teaching materials for qualitative methods. For example, the NSF could fund development of a multi-disciplinary annotated bibliography of foundational qualitative methods texts (note: one of the immediate results of the NSF conference that produced the present report is that a preliminary version of such a bibliography, though not annotated, is posted on the web site of the Consortium on Qualitative Research Methods at: http://www.asu.edu/clas/polisci/cqrm/syllabi.html

Grants and Fellowships

In addition to providing increased funding for such training initiatives, the NSF could also dedicate more resources to supporting actual qualitative research. Specific actions the NSF could take to increase the quantity and quality of qualitative research in political science and related disciplines include:

- Sponsoring research in a variety of areas of qualitative research (such as textual analysis, process tracing, and ethnographic work across disciplinary divides).
- Simplifying the procedures for small grants (under $20,000) and exploratory research, and simplify the application procedures for pre-dissertation pilot grants, and expand the availability of such grants.
- Developing programs and grants for training multi-method methodologists. Grants and training programs might enable qualitative researchers, including both graduate students and faculty, to develop skills in statistical methods or formal models, and vice-versa. Such grants could be targeted at PhD programs in political science that provide intensive training in statistical theory and qualitative methods, up to and including an MA in statistics. Funds might also go to workshops focused on the intersection of statistical theory and qualitative methods. NSF funds might also support faculty who are qualitative scholars and who undertake methodological training at ICPSR, and it might support faculty trained in statistical methods who attend the IQRM, and NSF funds might further enable and encourage these respective training programs to broaden their offerings on the intersections among fundamental statistical theory, qualitative research, and econometric methods. Just as the National Science Foundation's "Empirical Implications of Theoretical Models" (NSF-EITM) initiative has forged new connections between theoretical models and empirical analysis within the political science discipline, funding along these lines can help build a new bridge between statistical theory and qualitative research.

Political Science
Papers Presented by Workshop Participants

Overview of Organizational Efforts on Qualitative Methods in Political Science

Andrew Bennett
Department of Government
Georgetown University

Before turning to the workshop issues, I want to apprise scholars in other fields of several notable recent organizational efforts on behalf of qualitative methods in the field of political science. In the last five years, qualitative methods have attained a higher profile and more institutionalized role in political science that is more commensurate with their importance in published research in the field. One 2003 survey shows that nearly half the articles published in the top political science journals use qualitative methods, and that this proportion has remained fairly steady for the last two decades (Bennett, Barth, and Rutherford, 2003). Yet the same survey shows that qualitative methods are under-represented in graduate methodology courses, and until the last few years these methods lacked an institutional home comparable to the ICPSR (Inter-University Consortium for Political Science Research), which has successfully improved, disseminated, and promoted statistical research methods.

Several developments have begun to rectify these imbalances in the field. First, the formation of the Consortium on Qualitative Research Methods (CQRM) has created an institutional base for qualitative methods analogous to that which the ICPSR has provided for statistical methods. CQRM, headquartered at Arizona State University, has held four annual two-week training institutes in qualitative methods for advanced graduate students and junior faculty, training over two hundred students thus far. The institutes are held at ASU in the first two weeks of January. CQRM is funded by member departments and research centers, which pay membership dues in return for the opportunity to send students or junior faculty to the annual institute, and the consortium presently includes essentially every top political science department as a member.

Second, the creation of a new qualitative methods section in the American Political Science Association (APSA-QM) has given these methods a prominent role in APSA. After only two years of existence, APSA-QM has over 700 members, making it one of the largest APSA's three dozen sections, and it has an excellent newsletter now in its fourth issue. APSA-QM and CQRM jointly maintain a web site, at http://www.asu.edu/clas/polisci/cqrm/, which includes many syllabi and teaching materials in addition to information on the two organizations and the annual institute.

Third, for the past several years George Washington University has sponsored a four day summer institute on conducting archival research. The third such institute is planned for June, 2005, and will accept up to 25 students.

Finally, a group of European scholars has created an organization devoted to improving and disseminating qualitative methods through a series of conferences, workshops and related activities. This organization, known as COMPASSS (Comparative methods for the Advancement of Systematic cross-case analysis and Small-n Studies) maintains a web site that includes publications, working papers, software, and an extensive bibliography and list of scholars interested in qualitative methods, as well as announcements of ongoing COMPASSS activities.

Standards of Rigor For Qualitative Work in Political Science

While there is no "one size fits all" qualitative research design, and the many different kinds of research designs and comparisons (most similar cases, least similar cases, outlier or deviant cases, most- or least-likely cases, before-after cases, and so on) have different goals and requirements, there are a half-dozen general standards that help identify whether a piece of qualitative research has been done rigorously:

- **Case or site selection.** Has the researcher carefully selected the cases to be studied, and are these cases appropriate to the research design and objectives? Has the researcher defined "what this is a case of" and "what population are cases of it?" Has the researcher indicated what other cases were or could have been considered for close study, and why they were not chosen? Has the researcher considered what would constitute a "negative case" of the phenomenon, identified what cases might qualify as negative cases, and considered inclusion of such cases for purposes of comparison if this is appropriate to the research design and the phenomenon under study?
- **Conceptualization, operationalization, and measurement.** Has the researcher adequately conceptualized the phenomenon being studied, situated it with respect to the relevant theoretical or conceptual frameworks in the field, and defined the independent and dependent variables if the goal is causal inference? Are the measures of these concepts valid with respect to capturing the ideas or concepts they are meant to represent?
- **Attention to potential endogeneity.** Has the researcher considered and anticipated potential endogeneity, giving attention to possible feedback loops and selection effects that might create endogeneity or that might simply reverse the causal arrow from the direction hypothesized by the researcher?
- **Attention to the kind of case comparisons explicitly or implicitly used.** Depending on what kinds of comparisons are to be used, has the researcher adequately prepared and carried out these comparisons? For example, in a most-similar cases design, has the researcher adequately considered the residual differences between the cases to be compared, in addition to the one independent variable of interest that differs between the two cases, to be sure that these residual differences do not account for the difference in outcomes?
- **Attention to alternative explanations.** Has the researcher, in both cross-case comparisons and within-case process tracing, given fair and adequate attention to evidence for alternative explanations?
- **Careful empirical research, including into context, sequencing, coverage of a wide range of appropriate sources.** Has the researcher sufficiently immersed themselves in the details of the case(s) to make the kind of inferences to which they aspire? Have they researched the relevant primary sources thoroughly, as well as the secondary literature?

How Can Qualitative Methodologists in Political Science Better Share With and Learn From Other Fields?

While the organizational efforts on qualitative methods within political science have been substantial, these initiatives have not yet reached out to other fields. Opportunities for doing so include the following:

- **Publishing in journals in other fields, or co-authoring with those from other fields**. It would be very useful in this context to receive advice from scholars in other fields on which journals

in their fields would be most open to, and give the highest profile to, articles on qualitative methods. Within political science, the *American Political Science Review* has become much more receptive to qualitative work in recent years; other prominent journals open to such work include the journals *World Politics, International Organization, Comparative Politics,* and *Studies in Comparative International Development.*

- **Attending and making panel presentations at the associational conferences of other fields.** Information on which conferences are key, and when and where they are held, would be useful. The APSA conference is Labor Day weekend each year, and this year is in Washington D.C.; the 2006 conference will be in San Francisco.
- **Getting students and faculty from other fields to the CQRM Institute at ASU**. Thus far, sociologists have been the only scholars to attend other than political scientists. Departments in other fields are of course welcome to become members of the consortium and to send attendees to the institute. The institute also has about eight open slots for attendees from institutions that are not consortium members, but thus far these slots have gone mostly to political scientists.
- **Going international**. Joint efforts between among APSA-QM, CQRM, and COMPASSS have been limited thus far, but COMPASSS is working to get American scholars interested in qualitative methods to attend the International Political Science Association (IPSA) conference in Tokyo in the summer of 2006, and to hold a conference on qualitative methods in Tokyo immediately following the IPSA conference.
- **Sharing organizational email lists for announcements of mutual interest**. CQRM/APSA-QM make their mailing lists available for such purposes on a case-by-case basis.
- **Crossing disciplinary boundaries on PhD thesis committees**. My department has frequently included historians on our PhD committees.
- **Cross-listing relevant courses and getting them accepted into program requirements in other fields.**

What Areas or Topics Are Most Promising For Qualitative Methods?

In addition to reaching across disciplinary boundaries and building stronger ties internationally, I would suggest five priorities for future development of qualitative methods:

- **Improving techniques for and promoting the use of multi-method research.** There is tremendous interest in political science, especially among graduate students, in engaging in multi-method research. There is almost nothing written on how to combine research methods, however, and there are disciplinary/cultural barriers as well as technical challenges in doing so.
- **Improving methods for dealing with complexity**. There is great interest in complex relationships in political science - - path dependencies, tipping points, multiple interactions effects, and so on - - but these kinds of complexity pose difficult challenges for qualitative and other methods. Improved techniques for dealing with particular kinds of complexity, and exemplary empirical works on how to do so, are needed.
- **Developing more teaching materials for courses on qualitative methods**. In contrast to statistical and other methods, for which there are well-developed text books and numerous teaching materials like problem sets and exercises, there are relatively few teaching materials widely available for qualitative methods. The CQRM/APSA-QM web site includes some such materials, but much more needs to be developed and disseminated.

- **Studying new political phenomena**. While qualitative methods do not have a monopoly on studying new phenomena, they may be particularly useful at studying novel and under-theorized developments such as globalization, the political implications of electronic media, and terrorism in an age of weapons of mass destruction.
- **The study of American politics**. The survey noted above found that qualitative research on American politics has fallen from 12% of the articles in the top journals in 1975 to only 1% by 2000. A great opportunity exists for reviving a distinguished tradition of qualitative research on American politics.

References

Andrew Bennett, Aharon Barth, and Kenneth Rutherford. (2003). "Do We Preach What We Practice? A Survey of Methods in Journals and Graduate Curricula." *PS: Political Science and Politics* (July).

A Statistical Rationale for Qualitative Research?[12]

David Collier
Department of Political Science
University of California - Berkeley

In the spirit of Clifford Geertz's (1983: chapter.1) idea of "blurred genres," this document raises what might appear to be the paradoxical idea of a statistical rationale for qualitative research. More broadly, it considers the potential contribution of statistical theory—understood as a set of tools for reasoning about evidence and inference—to refining qualitative methods.[13] The document follows closely the outline provided for the NSF Workshop. This focus may be quite remote from the concerns of many workshop participants. Yet I am convinced that if we really are concerned with "interdisciplinary standards for systematic qualitative research," this is one of a number of promising avenues to pursue.

Identifying Standards of Rigor in Political Science

Political science is an incessant importer of methodologies, and it can easily be argued that political science methodology is sufficiently fractured as to make it misleading to speak of widely held standards of rigor. For present purposes, however, it is useful to comment on three potential sources of such standards.

Option 1. The Quantitative Template. This template is proposed in *Designing Social Inquiry* by King, Keohane, and Verba (1994),[14] who map the norms of large-N regression analysis onto qualitative research, seeking to provide standards and thereby offering the basis for "Scientific Inference in Qualitative Research," as they put it in the subtitle of their book. By contrast, in *Rethinking Social Inquiry: Diverse Tools, Shared Standards,* Brady and Collier (2004) and collaborators strongly dissent from this view, providing a comprehensive critique of the idea that the regression framework provides a general set of standards for guiding research. Regression analysis is unquestionably an indispensable analytic tool in the social sciences. Yet it is essential to not overestimate either the overall power of inference provided by regression-based research, or the power of inference contributed by specific tools and methodological injunctions associated with regression analysis and related quantitative techniques—for example, ideas about degrees of freedom, the traditional injunction concerning post-hoc hypothesis reformulation, and what can readily be seen as the careless extension of the idea of statistical significance.

Option 2. The Mainstream Framework of Qualitative Research. This well-established tradition in political science includes research techniques identified by such standard and quite familiar labels as the comparative method, comparative-historical method, qualitative-comparative analysis (QCA), concept-analysis as a branch of methodology, the case-study tradition, process tracing, pattern matching, typological theory, interpretivism, constructivism, and various traditions of field research—including, but not restricted to, ethnography. An extensive body of methodological work focused on these traditions strives for refinements, improvements, and greater rigor.[15]

12 The general framework for this discussion draws on Brady and Collier, eds. (2004), including five chapters co-authored with Jason Seawright. I take responsibility for the particular formulation of this framework presented here.

13 "Statistical theory" is thus seen as very different from conventional quantitative methods or the "quantitative template." For a discussion of four dimensions in terms of which it is useful to understand the quantitative-qualitative distinction (level of measurement, size of the N, whether statistical tests are employed, and thick versus thin analysis), see Collier, Brady, and Seawright (2004: 245-250).

14 See also the summary of King, Keohane, and Verba's argument in Collier, Seawright, and Munck (2004).

15 Examples of studies that seek to systematize these approaches include Adcock and Collier (2001), Collier (1993; but see also Hall 2004), Collier (unpublished manuscript), Collier and Adcock (1999), Collier and Levitsky (1997), Collier and Mahon (1993), Elman (2005), Finnemore and Sikkink (1999), George and Bennett (2005), Gerring (2001, 2004, 2005), Gerring and Barresi (2003), Gerring and Seawright (2005), Gerring and Thomas (2005), O'Brien (forthcoming), Mahoney and Rueschmeyer (2003), Ragin (1987, 2000) and Wedeen (2002, 2004).

Option 3. A Statistical Rationale for Qualitative Research. Perhaps paradoxically, "statistical theory," understood (as noted above) as a broad set of tools for evaluating evidence and inference, can provide a strong rationale for pursuing qualitative rather than quantitative research, and/or for combining the two traditions.[16] More than a few statisticians are skeptical about much of the regression/econometric tradition in the social sciences,[17] and some are convinced that fine-grained evidence about causal processes is an indispensable supplement to inferences based on regression estimation.[18] For some scholars, rather than a supplement, such evidence should be seen as an essential point of departure. Further, basic tools of statistical theory—for example, tools of probability theory and Bayesian inference—have provided valuable insights into how scholars can most effectively make inferences from qualitative data.[19]

Applying these Standards in Other Disciplines

These standards appear highly relevant for other disciplines in which scholars either: (a) seek new criteria and standards for systematizing qualitative tools, which is indeed the theme of this NSF workshop; and/or (b) agree with the need to supplement regression techniques with other sources of evidence.

Most Promising Substantive Topics for Qualitative Methods

One view might be that qualitative methods (which are often more inductive) are particularly well suited for domains for which large, standardized data sets are not available, or about which there is little prior knowledge, or in which change is so rapid that, in effect, prior knowledge is limited. Alternatively, in areas of research where large standardized data sets are available and which are not necessarily subject to rapid change, scholars may give a high priority to introducing qualitative tools because either (a) the application of the quantitative template in these domains yields inferences that need to be supported, supplemented, and possibly corrected by inferences based on qualitative analysis, or (b) quantitative research has focused on substantively narrow lines of analysis that need to be supplemented, broadened, and possibly superseded by qualitative approaches.

A Promising Avenue of Research on Qualitative Methods

To cite again some of the relevant studies, these tools have proved to be highly relevant in advancing debates on case selection in small-N qualitative research.[20] Relatedly, these tools have proved valuable in discussing the strengths and weaknesses of Qualitative Comparative Analysis (QCA), regression analysis, and traditional case studies.[21]

16 Collier, Brady, and Seawright (2004).

17 For a strong critique, from a statistical point of view, of many quantitative studies in the social sciences, see Freedman (2005: passim, e.g., Chapter 8, Section 8.11).

18 Goldthorpe (2001).

19 For example, Dion (1998); Ragin (2000: passim); Goertz and Starr (2001: passim); Seawright (2002a) and the accompanying symposium with contributions by Clarke (2002), Braumoeller and Goertz (2002), and a response by Seawright (2002b); and the forthcoming symposium in *Studies in Comparative International Development*, with contributions by Seawright, Achen, and Ragin, and a response by Seawright.

20 Dion (1998), Ragin (2000), Seawright (2003), and Goertz and Starr (2003).

21 See Seawright (forthcoming), Achen (forthcoming), and Seawright (forthcoming, response).

References

Achen, Christopher H. (forthcoming, 2005). "Two Cheers for Charles Ragin." *Studies in Comparative International Development 40,* No. 1 (Summer).

Adcock, Robert, and David Collier. (2001). "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95: 529-46.

Brady, Henry, and David Collier, eds. (2004). *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield.

Braumoeller, Bear, and Gary Goertz. (2002). "Watching Your Posterior: Bayes, Sampling Assumptions, Falsification, and Necessary Conditions." *Political Analysis 10*: 198-203.

Clarke, Kevin A. (2002). "The Reverend and the Ravens: Comment on Seawright." *Political Analysis* 10: 194-97.

Collier, David. (2005). "Getting Started on a Research Design: Launching Qualitative Research Projects." Unpublished manuscript, Department of Political Science, University of California, Berkeley.

Collier, David. (1993). "The Comparative Method." In Ada W. Finifter, ed., *Political Science: The State of the Discipline II.* Washington, D.C.: American Political Science Association.

Collier, David, and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts." *Annual Review of Political Science* Vol. 2. Palo Alto: Annual Reviews. Pp. 537-65.

Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." Chapter 13 in Henry E. Brady and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield.

Collier, David and Steven Levitsky. 1997. "Democracy with Adjectives: Conceptual Innovation in Comparative Research." *World Politics 49* (April): 430-51.

Collier, David, and James E. Mahon. 1993. "Conceptual `Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87, No. 4: 845-55.

Collier, David, Jason Seawright, and Gerardo L. Munck. 2004. "The Quest for Standards: King, Keohane, and Verba's Designing Social Inquiry." In Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield. Pp. 21-50.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30, No. 2 (January): 127-45.

Elman, Colin. 2005. "Explanatory Typologies in Qualitative Studies of International Politics." *International Organization* 59, No. 2 (Spring): 293-326.

Finnemore, Martha, and Kathryn Sikkink. 1999. "Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics." *Annual Review of Political Science* Vol. 4. Palo Alto: Annual Reviews. Pp. 391-416.

Freedman, David. 2005. *Statistical Models: Theory and Practice.* New York: Cambridge University Press.

Explanatory Typologies in Qualitative Analysis

Colin Elman
Department of Political Science
Arizona State University

Typologies have a distinguished history in the social and natural sciences, and discussions of what they are and how they work have generated a large body of literature.[22] This memo focuses on what I will call explanatory typologies, by which I mean multidimensional conceptual classifications based on an explicitly stated theory.[23]

Explanatory typologies are likely to be most valuable for qualitative analysis when scholars systematically apply shared techniques.[24] Drawing on a recent article in *International Organization*,[25] this memo provides a brief account of steps used in working with typologies, and an accessible vocabulary to describe them. Two groups of techniques are of particular interest when refining typologies: compression and expansion.[26]

Compression facilitates working with multivariable explanatory typologies that would otherwise be too large and complex to be helpful. Five forms of cell compression are considered: *rescaling compression* (reducing the level of measurement); *indexing* (treating equal totals of additive causal variables as equivalent); *logical compression* (deleting cells that are the product of impossible or highly improbable combinations of variables); *empirical compression* (deleting empty cells); and *pragmatic compression* (collapsing contiguous cells if their division serves no useful theoretical purpose). The expansion of a partial typology allows for the rediscovery of deleted cells. This permits the analyst to discover missed combinations and suppressed assumptions, and to identify important cases.

Explanatory Typologies

Explanatory typologies invoke both the descriptive and classificatory roles of typologies albeit, as noted in Table 1, in a way that incorporates their theoretical focus. At its most straightforward, the descriptive role builds types from the "compounds of attributes" of concepts.[27] Each unique combination of the attributes of the included concepts provides a separate compound concept. Conventional usage arrays the component attributes in rows and columns to construct an associated property space. Every cell in that space captures a possible grouping of the attributes of the concepts being organized.[28]

In an explanatory typology, the descriptive function follows the conventional usage, but in a way that is heavily modified by its theoretical purposes. The constituent attributes are extracted from the variables of a preexisting theory. The dimensions of the property space (its rows and columns)

22 For overviews and reviews see, for example, Capecchi 1968; Nowotny 1971; Marradi 1990; Bailey 1972, 1973, 1992, and 1994; and Tiryakian 1968. Mastering this literature is made difficult by the proliferation of labels for different kinds of types, including extreme, polar, ideal, pure, empirical, classificatory, constructed, and heuristic. In addition, methodologists tend to invent new terms for the different components in their "typology of typologies," and to then compare their approach with previous treatments of other scholars. As a result, the choice of labels for describing the subset of typologies and typological procedures discussed in this memo is somewhat arbitrary.

23 The approach to typologies taken partly parallels J.W.N. Watkins' (1953) reading of Max Weber's "individualistic" ideal types. (See also McIntosh 1977: 267, n. 11; and Lindbekk 1992: 292-295. For different interpretations of ideal types see Albrow 1990; Burger 1987: 160-167; 2001; Clarke 2001; Hekman 1983; and Rogers 1969.) It is also sympathetic to, but goes beyond, Arthur Stinchecombe's (1987: 43-47) description of type-concepts and typologies. The approach is also consistent with that taken by Charles Ragin (2000: 76-87) but without adopting his Boolean data analytic strategy of qualitative comparative analysis (QCA) or fuzzy sets (ibid., 120-145; Ragin 1987).

24 Political science has benefited from several recent innovative treatments of qualitative methods, including King, Keohane and Verba 1994; Gerring 2001; Brady and Collier 2004; and George and Bennett 2005.

25 Elman 2005. The article from which this memo was drawn benefited greatly from extensive critiques of several drafts by David Collier. Stephen G. Walker, Miriam Fendius Elman, James Mahoney, Gary Goertz, Reilly O'Neal, John Gerring, Bear Braumoeller, Lisa Martin, two anonymous reviewers, and the participants at the January 2004 Institute for Qualitative Research Methods at Arizona State University provided valuable comments.

26 Lazarsfeld 1937; Lazarsfeld and Barton 1965; and Barton 1955 prefer the labels "reduction" and "substruction" to compression and expansion.

27 Lazarsfeld 1937, 120.

28 Lazarsfeld and Barton 1965, 169.

reflect alternative values of the theory's independent variables, so each cell in the space is associated with predicted values of the theory's intervening or dependent variables.[29] This association changes the descriptive question being answered from "What constitutes this type?" to "If my theory is correct, what do I expect to see?"

The classificatory function of typologies determines to which "type" a case can be characterized as belonging. Beginning with a typology, empirical data is coded as falling into one cell or another, guiding scholars to answer the question "What is this a case of?" The property space can be used to map, and compare, a population of cases by their respective scores on the component attributes of the typology.

In explanatory typologies, the classificatory function focuses exclusively on evidence that can arbitrate the theoretical claims being made. For example, analysts may investigate a case to determine whether there is the anticipated congruence between its scores on the typology's dimensions, and the predictions made in the cell in which the case is expected to belong. In addition, the analyst can use the location of cases in different cells as a guide to making the most productive comparisons for testing the underlying theory.

An explanatory typology is based on an explicitly stated preexisting theory. That theory may have been originally derived inductively from observations, or deductively using ordinary language or formal methods. Regardless of how the theory was first produced, however, an explanatory typology is primarily a complement to deductive approaches, because filling in the cells requires working through the logical implications of the theory: given its posited causal relationships, what particular outcomes are associated with different combinations of values of the theory's variables? The dimensions of the property space are provided by the theory's explanatory variables, and the content of the cells come from the logic of the theory.

Goals of Typologies

	Descriptive	**Classificatory**	**Explanatory**
Analytic Move(s)	Defines compound concepts (types) to use as descriptive characterizations.	Assigns cases to types.	Makes predictions based on combinations of different values of a theory's variables. Places data in relevant cells for congruence testing and comparisons to determine whether data is consistent with the theory.
Question(s) answered	What constitutes this type?	What is this a case of?	If my theory is correct, what do I expect to see? Do I see it?
Example	What is a parliamentary democracy as opposed to a presidential democracy?	Are Britain and Germany parliamentary or presidential democracies?	According to the normative variant of the democratic peace theory, what foreign policy behavior is predicted from a dyad of two mature parliamentary democracies? Do the bilateral foreign policies of Britain and Germany agree with that prediction?

29 See McKinney 1950, 238 and 1954, 164-169 on the relationship between theories and typologies. Note, however, that McKinney's "constructive typologies" are not always, or perhaps not even usually, theoretical in the sense used in this article. For example, while McKinney (1954, 195; 1966, 63) acknowledges that typologies can be derived from theories, he also suggests (1966, 63) that they can most usefully be constructed directly from the particularities of a historical situation.

The focus in this memo is on how explanatory typologies can be helpful to qualitative scholars, who have traditionally combined language theorizing with the intensive study of a small number of cases using comparative ordinary case, process-tracing, and congruence-testing methods. Qualitative scholars can enhance both the development and testing of their theories with a more self-conscious application of typological procedures. With respect to theory development, typologies are complementary to specifiying configurative or conjunctive causation,[30] describing equifinality or multiple sufficient causation,[31] and building in temporal effects.[32] With respect to theory testing, typologies help scholars to identify the degree of casual homogeneity between cells,[33] and to engage in counterfactual reasoning.[34]

Explanatory typologies are likely to be most valuable in the qualitative study of international politics when scholars self-consciously employ typological techniques. To date, the international relations subfield has lacked an account of what analytic moves are available, an accessible vocabulary to describe them, and concrete examples of how these techniques can be applied. The next two sections of this memo build on Lazarsfeld and Barton to develop procedures for manipulating explanatory typologies, looking first at techniques for compression, and then for expansion.

Compressing the Number of Cells in a Property Space

Lazarsfeld (1937), building on Hempel and Oppenheim (1936), provided the seminal discussion of the different techniques for compressing a property space, later developing them further with Allen Barton.[35] Lazarsfeld and Barton (1965, 173) define a "reduction" (what I am calling a "compression) as "any classification as a result of which different combinations fall into one class." Compression procedures include:

Rescaling. The number of cells can be reduced by lowering the number of attributes for one or more of the theory's variables represented in the property space. For example, changing a four variable model from trichotomous to dichotomous measurement reduces the number of cells from eighty-one to sixteen.[36] One thing to keep in mind when reducing the number of attributes is that each cell in the typology becomes more inclusive, hence potentially grouping cases which may not fit comfortably together.[37]

Indexing. Barton (1955, 46) observes that where multiple attributes "express essentially the same underlying characteristic or have their effects in the same direction" we can "give each category on each dimension a certain weight, and add these together to get index scores for each cell." Indexing treats all combinations that receive the same score as equivalent, in effect "folding over the typology thus rendering formerly distant types equal."[38] This technique presents more complex difficulties than its seeming simplicity would suggest, requiring arbitrary decisions on the appropriate weight for the high-mid-low rank on each attribute. It should also be noted that indexing presupposes that equal scores are equivalent. It may be that interaction effects between the different variables render this assumption problematic.

Logical Compression. There may be a connection between two or more of the typology's

30 Ragin 2000, 67—82. See Brady 2002 for an outstanding review of different models of causal inference.

31 Bennett 1999a, 9. See also Bennett and George 2001, 138.

32 On such phenomena, see Pierson 2000, 2003 and 2004; Mahoney 2000; Buthe 2002; Thelen 2003; and Aminzade 1992. On typologies and time, see Nowotny 1971.

33 See Munck 2004, 111; Nowotny 1971, 6—11; Rogowski 2004, 7; McKeown 2004, 13; Eckstein 1975, 117—20; and Przeworski and Teune 1970, 32—39.

34 Tetlock and Belkin 1996, 4. See also Fearon 1991; and Hempel 1965b, 164—65.

35 Lazarsfeld and Barton 1965; Barton 1955.

36 On levels of measurement see Stevens 1946, 1951 23-30, and 1959.

37 For discussions of the closely related issues of conceptual differentiation and stretching, see Sartori 1970; Collier and Mahon 1993; Collier and Levitsky 1997; Collier and Adcock 1999; Sartori 1984; and Gerring 1999. On the connection between concepts and classification, see Hempel 1965a, 138—139, 146—148.

38 Bailey 1994, 28.

dimensions such that some combinations are logically impossible or highly improbable. If so, we can delete these cells.[39] It should be noted that logical compression is a characteristic of the underlying theory, and it will almost always be an option to add an auxiliary assumption that will render an otherwise unfeasible prediction possible.

Empirical Compression. Some combinations of variables may be logically possible or not highly improbable, but there may nevertheless be no empirical examples of those combinations. If so, we may be able to delete these cells from the typology.[40]

Pragmatic Compression. Scholars can collapse contiguous cells if their division serves no useful theoretical purpose. Using pragmatic compression, "certain groups of combinations are contracted to one class in view of the research purpose."[41]

Expanding Property Space

Explanatory typologies can be constructed directly from a theoretical statement. They can also be rebuilt from analyses that already use a typology which has previously been minimized. This section considers the technique of expanding a property space from such a partial typology.

Expansion (what Lazarsfeld (1937: 132) calls "substruction") takes an underspecified typology, or one that is implied from the use of a sub-population of its types, and provides a full account of the associated property space by 'reverse engineering' the classification. The analyst works backwards to lay out the property space from which the partial typology is derived, and the type of reduction technique that was used to produce it. As Lazarsfeld (1937: 132) notes, the procedure does not assume "that the creator of the types really had such a procedure in mind. It is only claimed that, no matter how he actually found the types, he could have found them logically by such an expansion."

Typological expansion allows analysts to spot important combinations of attributes that were overlooked in the partial typology, and to draw attention to cases that need further attention.[42] The procedure may also help theorists to make explicit the assumptions that were used by the original analyst to suppress particular combinations. [43] The technique can be used to draw out the implications of a theorist employing outstanding "types" with different attribute clusters, or to expand a formal but reduced typology back to its complete specification.

Pitfalls in Property Space: Reification and Puzzle Relabeling

The cells in an explanatory typology are best seen as "containers" of predictions made by the underlying theory. Users of explanatory typologies have to avoid a form of reification,[44] where cell labels themselves become free-standing "explanations," rather than the theory from which the property space was derived. To put it another way, in the context of an explanatory typology, reification occurs when a case is "explained" because we attach a name to it, not because a theory we have deemed valid is seen as being applicable to it. This is less likely to be a problem for the original developer of an explanatory typology, but may well be an issue for scholars who read and use the typology at one remove.

39 Lazarsfeld 1937, 126; Lazarsfeld and Barton 1965, 173.

40 Bailey 1994, 27; Barton 1955, 46, 49; and Marradi 1990, 144.

41 Lazarsfeld and Barton 1965, 174. See also Lazarsfeld 1937, 128; Bailey 1994, 27; Barton 1955, 45-46; and Marradi 1990, 144.

42 Barton 1955, 53.

43 Barton 1955, 50.

44 On the dangers of reification see Bailey 1994, 15; Tiryakian 1968, 179; McKinney 1954, 148-149.

A second challenge is whether a typology is really explanatory, or is instead a form of semantic relabeling which displaces questions without really answering them.[45] Although framed here in the context of increasing the number of cells in a property space, this issue arises whenever theories are amended to cover known anomalies.[46] Philosophers of science worry that an amendment to a theory designed to address a puzzle may just be a move to protect it from falsification, and not real scientific progress.[47] Typically, a concept is redefined, or an auxiliary hypothesis is added, to allow the theory to predict the anomaly.

One way of addressing this problem is to adopt a form of what Alan Musgrave (1974: 3, 7) calls the historical approach to confirmation of a scientific theory.[48] The historical approach suggests that we cannot determine whether evidence supports a theory solely on the basis of whether it "fits" the current iteration of the theory. It is not enough to ask whether the theory covers known anomalies. It is also necessary to track the trajectory of a theory as it develops, and ask whether amendments did more than just relabel empirical puzzles. The question would be whether the new categories provide additional value, signaled by the prediction of novel facts.[49]

Conclusion

While political scientists commonly employ explanatory typologies in their analysis of international politics, the sub-field has not paid sufficient attention to the logic that underlies and justifies that usage, or to the different techniques that are available for expanding and compressing property space. While the procedures described in this memo may seem habitual, even intuitive, explanatory typologies are at their most powerful when they are used self-consciously. A more grounded approach will encourage rigor, enhance transparency, and increase the likelihood of producing cumulative results.

45 Vasquez 1997 makes a similar critique.

46 This kind of 'iteration' between theory and evidence is often prescribed (see, for example, Bates et. al. 1998, 16; and Morrow 2002, 187-188) and hence the problem is likely to arise often.

47 The best known discussion of this issue is Lakatos 1970.

48 See also Worrall 1978b, 321; and Mayo 1996, 254-256. It should be noted that the historical approach to confirmation looks for different categories of predictions, not evidence that predicted values of the same dependent variable are repeated in additional cases. To be sure, recurring tests of the same proposition are valuable because they offer evidence about whether a prediction is empirically accurate. As Robert Jervis (1985, 146) notes, "Scholars often look at many cases to see if a proposed generalization fits the data. [But t]his is a form of confirmation, not the discovery of new facts."

49 Philosophers of science disagree on which standard of novelty to apply, i.e. they differ on the answer to the question: "novel compared to what?" Potential answers to that question include: strict temporal novelty (Lakatos 1970, 118; Zahar 1973, 101; Worrall 1978a, 46, and 66, n. 7; Frankel 1979, 24; Gardner 1982, 2; Nunan 1984, 275; and Hands 1991, 96); new interpretation novelty (Lakatos 1970, 188; Koertge 1971, 171, n. 5; Musgrave 1974, 11; Nunan 1984, 275; Carrier 1988, 207); heuristic novelty (Zahar 1973, 101; Lakatos and Zahar 1975, 376, n. 65); and background theory novelty (Musgrave 1974, 15-16; Worral 1978b, 321-322; Mayo 1996, 208).

References

Albrow, Martin. (1990). *Max Weber's Construction of Social Theory*. New York: St. Martin's Press.

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

Bailey, Kenneth D. (1972). "Polythetic Reduction of Monothetic Property Space." *Sociological Methodology,* 4:83-111.

Bailey, Kenneth D. (1973). "Monothetic and polythetic typologies and their relation to conceptualization measurement, and scaling." *American Sociological Review,* 38:18-33.

Bailey, Kenneth D. (1992). "Typologies." Pp. 2188-2194 in *Encyclopedia of Sociology*, edited by E. F. Borgatta and M. L. Borgatta. New York: Macmillan.

Bailey, Kenneth D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks: Sage Publications.

Barton, A. H. (1955). "The Concept of Property-Space in Social Research." Pp. 40-53 in *Language of Social Research*, edited by P. F. Lazarsfeld and M. Rosenberg. Glencoe: Free Press.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. (1998). *Analytic Narratives*. Princeton: Princeton University Press.

Bennett, Andrew. (1999). "Causal Inference in Case Studies: From Mill's Methods to Causal Mechanisms." Paper presented at the 96th Annual Meeting of the American Political Science Association, September, Atlanta, Georgia.

Bennett, Andrew and Alexander George. (2001). "Case Studies and Process Tracing in History and Political Science: Similar Strokes for Different Foci." Pp. 137-166 in *Bridges and Boundaries: Historians, Political Scientists and the Study of International Relations*, edited by C. Elman and M.F. Elman. Cambridge: MIT Press.

Brady, Henry E. (2002). "Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland Theory." Paper presented at the 16th Annual Meeting of the Political Methodology Group, July, Seattle, Washington.

Brady, Henry and David Collier. (2004). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham: Rowman and Littlefield.

Burger, Thomas. (1987). *Max Weber's Theory of Concept Formation*. Durham: Duke University Press.

Buthe, Tim. (2002). "Taking Temporality Seriously: Modeling History and the Use of Narrative as Evidence." *American Political Science Review,* 96: 481-493.

Capecchi, Vittorio. (1968). "On the Definition of Typology and Classification in Sociology." *Quality and Quantity,* 2: 9-30.

Carrier, Martin. (1988). "On Novel Facts: A Discussion of Criteria for Non-ad-hoc-ness in the Methodology of Scientific Research Programmes." *Zeitschrift fur allgemaine Wissenschaftstheorie,* 19: 205-231.

Clarke, S. (2001). "Idealization, Abstraction and Ideal Types." Pp. 7142-7148 in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Bates. New York: Elsevier.

Collier, David and Robert Adcock. (1999). “Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts.” *Annual Review of Political Science,* 2: 537-65.

Collier, David and Steven Levitsky. (1997). “Democracy with Adjectives: Conceptual Innovation in Comparative Research.” *World Politics,* 49: 430-451.

Collier David, and James E. Mahon. (1993). “Conceptual ‘Stretching’ Revisited: Adapting Categories in Comparative Analysis.” *American Political Science Review,* 87: 845-855.

Eckstein, Harry. (1975). “Case Study and Theory in Political Science.” Pp. 79-137 in *Handbook of Political Science,* edited by F. Greenstein and N. Polsby. Reading: Addison-Wesley

Elman, Colin. (2005). “Explanatory Typologies in Qualitative Studies of International Politics.” *International Organization,* 59: 293-326.

Fearon, James D. (1991). “Counterfactuals and Hypothesis Testing in Political Science.” *World Politics,* 43:169-195.

Frankel, Henry. (1979). “The Career of Continental Drift Theory: An Application of Imre Lakatos’ Analysis of Scientific Growth to the Rise of Drift Theory.” *Studies in History and Philosophy of Science,* 10: 21-66.

Gardner, Michael R. (1982). “Predicting Novel Facts.” *British Journal for the Philosophy of Science,* 33: 1-15.

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

Gerring, John. (1999). “What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences.” *Polity,* 31: 357-393.

Gerring, John. (2001). *Social Science Methodology: A Criterial Framework*. Cambridge: Cambridge University Press.

Hands, D. Wade. (1991). “Reply to Hamminga and Maki.” Pp. 91-102 in *Appraising Economic Theories: Studies in the Methodology of Research Programs*, edited by N. de Marchi and M. Blaug. Brookfield: Edward Elgar.

Hempel, Carl G. (1965a). “Fundamentals of Taxonomy.” Pp. 137-154 in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, edited by C. G. Hempel. New York: The Free Press.

Hempel, Carl G. (1965b). “Typological Methods in the Natural and Social Sciences.” Pp. 155-171 in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, edited by C. G. Hempel. New York: The Free Press.

Hempel, Carl. G. and P. Oppenheim. (1936). *Der Typusbegriff im Lichte der Nuen Logik*. Leiden, Netherlands: A. W, Siythoff.

Hekman, Susan J. (1983). “Weber’s Ideal Type: A Contemporary Reassessment.” *Polity,* 16: 119-137.

Jervis, Robert. (1985). “Pluralistic Rigor: A Comment on Bueno de Mesquita.” *International Studies Quarterly,* 29: 145-149.

King, Gary, Robert O. Keohane, and Sidney Verba. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Koertge, Noretta. (1971). “Inter-Theoretic Criticism and the Growth of Science.” Pp. 160-173 in *Boston Studies in the Philosophy of Science, Volume 8: PSA 1970*, edited by R.C. and R. S. Cohen Buck. Dordrecht, Netherlands: D. Reidel

Lakatos, Imre. (1970). “Falsification and the Methodology of Scientific Research Programmes. Pp. 91-196 in *Criticism and the Growth of Knowledge,* edited by I. Lakatos and A. Musgrave. New York: Cambridge University Press.

Lakatos, Imre and Elie Zahar. (1975). “Why Did Copernicus’ Research Programme Supersede Ptolemy’s?” Pp. 354-383 in *The Copernican Achievement*, edited by Robert S. Westman. Berkeley: University of California Press.

Lazarsfeld, P.F. (1937). “Some Remarks on the Typological Procedures in Social Research.” *Zeitschrift fur Sozialforschung,* 6: 119-139.

Lazarsfeld, Paul F and Allen H. Barton. (1965). “Qualitative Measurement in the Social Sciences: Classification, Typologies, and Indices.” Pp. 155-192 in *The Policy Sciences*, edited by D. Lerner and H. D. Lasswell. Stanford: Stanford University Press.

Lindbekk, Tore. (1992). “The Weberian Ideal-Type: Development and Continuities.” *Acta Sociologica,* 35: 285-297.

Mahoney, James. (2000). “Path Dependence in Historical Sociology.” *Theory and Society,* 29: 507-548.

Mahoney, James and Dietrich Rueschemeyer, eds. (2003). *Comparative-Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Marradi, Alberto. (1990). “Classification, Typology, Taxonomy.” *Quality and Quantity,* 24: 129-157.

Mayo, Deborah G. (1996). *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

McIntosh, Donald. (1977). “The Objective Bases of Max Weber’s Ideal Types.” *History and Theory,* 16: 265-279.

McKeown, Timothy J. (2004). “Case Studies and the Limits of the Statistical Worldview.” Pp. 139-167 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by H. E. Brady and D. Collier. Lanham: Rowman and Littlefield.

McKinney, John C. (1950). “The Role of Constructive Typology in Scientific Sociological Analysis.” *Social Forces,* 28: 235-240.

McKinney, John C. (1954). “Constructive Typology and Social Research.” Pp. 139-198 in *An Introduction to Social Research*, edited by J. T. Doby. Harrisburg: The Stackpole Company.

McKinney, John C. (1957). “The Polar Variables of Type Construction.” *Social Forces,* 35: 300-306.

McKinney, John C. (1966). *Constructive Typology and Social Theory*. New York: Appleton-Century-Crofts.

McKinney, John C. (1969). “Typification, Typologies and Sociological Theory.” *Social Forces,* 48: 1-12.

Mearsheimer, John J. (2001). *The Tragedy of Great Power Politics*. New York: Norton and Company.

Morrow, James D. (2002). “International Conflict: Assessing the Democratic Peace and Offense-Defense Theory.” Pp. 172-196 in *Political Science: State of the Discipline,* edited by I. Katznelson and H.V. Milner. New York: W.W. Norton & Co.

Munck, Gerardo. (2004). “Tools for Qualitative Research.” Pp. 105-121 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards,* edited by H. E. Brady and D. Collier. Lanham: Rowman and Littlefield.

Musgrave, Alan. (1974). “Logical versus Historical Theories of Confirmation.” *British Journal for the Philosophy of Science,* 25: 1-23.

Nowotny, Helga. (1971). “The Uses of Typological Procedures in Qualitative Macrosociological Studies.” *Quality and Quantity,* 6: 3-37.

Nunan, Richard. (1984). “Novel Facts, Bayesian Rationality, and the History of Continental Drift.” *Studies in History and Philosophy of Science*, 15: 267-307.

Pierson, Paul. (2000). “Increasing Returns, Path Dependence, and the Study of Politics.” *American Political Science* Review, 94: 251-267.

Pierson, Paul. (2003). “Big, Slow-Moving, and . . . Invisible: Macro-Social Processes in the Study of Comparative Politics.” Pp. 177-207 in *Comparative Historical Analysis in the Social Sciences,* edited by J. Mahoney and D. Rueschemeyer. Cambridge: Cambridge University Press.

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

Przeworksi Adam and Henry Teune. (1970). *The Logic of Comparative Social Inquiry*. New York: Wiley Interscience.

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

Ragin, Charles. (2000). *Fuzzy Set Social Science*. Chicago: University of Chicago Press.

Rogers, Rolf E. (1969). *Max Weber’s Ideal Type Theory*. New York: Philosophical Library, Inc.

Rogowski, Ronald. (2004). “How Inference in the Social (but not the Physical) Sciences Neglects Theoretical Anomaly.” Pp. 75-83 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by H. E. Brady and D. Collier. Lanham: Rowman and Littlefield.

Sartori, Giovanni. (1970). “Concept Misformation in Comparative Politics.” *American Political Science Review,* 64: 1033-1053.

Sartori, Giovanni. (1984). “Guidelines for Concept Analysis.” Pp. 15-85 in *Social Science Concepts: A Systematic Analysis*, edited by Giovanni Sartori. Beverly Hills: Sage.

Schweller, Randall L. (1998). *Deadly Imbalances: Tripolarity and Hitler’s Strategy of World Conquest*. New York: Columbia University Press.

Stevens, S.S. (1946). “On the Theory of Scales of Measurement.” *Science*, 103: 677-680.

Stevens, S.S. (1951). “Mathematics, Measurement, and Psychophysics.” Pp. 1-49 in *Handbook of Experimental Psychology*, edited by S.S. Stevens. New York: John Wiley and Sons, Inc.

Stevens, S.S. (1959). “Measurement, Psychophysics, and Utility.” Pp. 18-63 in *Measurement: Definitions and Theories*, edited by C. W. Churchman and P. Ratoosh. New York: John Wiley and Sons, Inc.

Stinchecombe, Arthur L. (1987). *Constructing Social Theories*. Reprint, originally published 1968 by Harcourt, Brace & World. Chicago: University of Chicago Press.

Tetlock Philip E. and Aaron Belkin. (1996). "Counterfactual Thought Experiments in World Politics: Logical, Methodological and Psychological Perspectives." Pp. 1-38 in *Counterfactual Thought Experiments in World Politics*, edited by P. E. Tetlock and A. Belkin. Princeton: Princeton University Press.

Thelen, Kathleen. (2003). "How Institutions Evolve: Insights from Comparative-Historical Analysis." Pp. 208-240 in *Comparative-Historical Analysis in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. Cambridge: Cambridge University Press.

Tiryakian, Edward A. (1968). "Typologies." Pp. 177-186 in *International Encyclopedia of the Social Sciences*, edited by D. L. Sills. New York: Macmillan Co.

Vasquez, John A. (1997). "The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz's Balancing Proposition." *American Political Science Review,* 91: 899-912.

Walt, Stephen M. (1987). *The Origins of Alliances.* Ithaca: Cornell University Press.

Watkins, J.W.N. (1953). "Ideal Types and Historical Explanation." Pp. 723-743 in *Readings in the Philosophy of Science*, edited by H. Feigl and M. Brodbeck. New York: Appleton-Century-Crofts Inc.

Worrall, John. (1978a). "The Ways in Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology." Pp. 45-70 in *Progress and Rationality in Science,* edited by G. Radnitzky and G. Andersson. Dordrecht, Netherlands: D. Reidel.

Worrall, John. (1978b). "Research Programmes, Empirical Support and the Duhem Problem: Replies to Criticism." Pp. 321-338 in *Progress and Rationality in Science*, edited by G. Radnitzky and G. Andersson. Dordrecht, Netherlands: D. Reidel.

Zahar, Elie. (1973). "Why Did Einstein's Programme Supersede Lorentz's?" *British Journal for the Philosophy of Science*, 24: 95-123.

What Standards Are (or Might be) Shared?

John Gerring
Department of Political Science
Boston University

For some time, methods have been associated with quantitative methods. Thankfully, this has begun to change. Yet, there is still very little space for non-quantitative styles of research in the discipline of political science – relative, that is, to what is available for quantitative research. Witness: the contrast between "Arizona" (IQRM), with its annual contingent of 80+ students, and "Michigan" (ICPSR), with its hundreds of annual participants. Witness: the relative paucity of non-quantitative methods courses at the graduate level. Witness: the ongoing resistance on the part of established journals (e.g., *American Journal of Political Science, Political Analysis*) to the inclusion of work by qualitative methodologists. Witness: the general confusion and consternation about what constitutes a solid, methodologically defensible, qualitative study.

My first recommendations are therefore quite simple, though rather difficult to implement: greater attention should be paid to the qualitative aspects of political science methods. This should involve the expansion of IQRM/CQRM, the creation of new (required) courses in graduate programs, greater openness on the part of editors and reviewers to qualitative methods, and more explicit care to these matters on the part of researchers in the field.

Yet, all of this presumes an important precondition: that a field can be created, or is in the process of creation, which is sufficiently explicit about methodological criteria and where sufficient consensus exists such that these norms – whatever they may be – can be taught, understood, and respected. This, I take it, is the central goal of NSF's current initiative.

Where shall we look for these cross-disciplinary, cross-subfield criteria? Here, I provide a brief and necessarily schematic treatment of arguments pursued at length elsewhere (see References).

Towards Common Criteria for Social Science Work

I have argued that the work of social science is usefully divided into three inter-dependent tasks: *concept formation*, *propositions*, and *research design*. Each of these tasks responds to a somewhat different set of demands. Thus, the vast and complex subject of social science methodology may be conceptualized as a set of discrete tasks and their attendant criteria.

It must be stressed that the following criteria are understood as general goals, not as necessary conditions. They are always applicable, but not always fully achievable. Indeed, the process of conducting research usually involves *tradeoffs* among these three tasks and their attendant criteria. It should also be noted that this framework explicitly excludes research issues pertaining to practical or logistical issues – e.g., funding, time, expertise, availability of data, and so forth. Practical matters are important, to be sure; but they are not methodological issues per se.

Concepts. Concepts answer the *what?* question. In order to talk about anything at all one must call it by a name. Since some names are better than others, and some definitions better than others, we cannot escape the problem of concept formation. Adequacy in concept formation obliges one to consider eight criteria more or less simultaneously: 1) *coherence*, 2) *operationalization*, 3) *validity*, 4) *field utility*, 5) *resonance*, 6) *contextual range*, 7) *parsimony*, and 8) *analytic/empirical utility*. Juggling these criteria successfully is the art of forming good concepts.

Propositions. Propositions involve the formulation of empirical statements about the phenomenal world. (Arguments, hypotheses, explanations, and inferences are all 'propositions' in the broad sense that I employ this term.) Propositions can be classified as descriptive, predictive, or causal. Causal propositions – the most complex, methodologically speaking – are subject to the following criteria: 1) *specification* (clarification of the range of outcomes under investigation, the set of cases to which the proposition refers, the resolution of any internal contradictions, and the operationalization of all key terms), 2) *precision*, 3) *breadth* (i.e., scope, range, and generality), 4) *boundedness* (the establishment of a logical and theoretically defensible set of boundaries for the proposition; that which it covers and that which it does not), 5) *completeness* (the degree of variance explained by the proposition), 6) *parsimony*, 7) *differentiation* (is X differentiable from Y), 8) *priority* (the causal distance between X and Y), 9) *independence* (the extent to which X is exogenous relative to Y), 10) *contingency* (the identification of a causal factor that is contingent, relative to what may be considered the normal course of events), 11) *mechanism* (the causal path connecting X and Y), 12) *analytic utility* (the extent to which a proposition accords with what we know about the world, including commonsense and theoretical knowledge), 13) *intelligibility*, 14) *relevance* (societal significance), 15) *innovation* (novelty), and 16) *comparison* (is the favored X better – along these various dimensions – than other possible Xs?). A good causal argument is well-specified, precise, broad, bounded, and so forth. (Descriptive and predictive proposition can also be understood in terms of these general criteria, though not all of these sixteen dimensions apply, or they apply somewhat differently.)

Research Designs in Causal Inference. The fundamental problem of causal inference is that we cannot re-run history to see what effects X actually had on Y in a particular case. At an ontological level, this problem is unsolvable. However, we have various ways of reducing this uncertainty such that causal inference becomes possible and plausible.

There are two dimensions upon which causal effects may be observed, the temporal and the spatial. Temporal effects may be observed directly when an intervention occurs: X intervenes upon Y and we observe any change in Y that may follow. Here, the "control" is the pre-intervention state of Y; what Y was prior to the intervention (a state that we presume would remain the same, or whose trend would remain constant, in the absence of an intervention). Spatial effects may be observed directly when two phenomena are similar enough to be understood as examples (cases) of the same thing. Ideally, they are similar in all respects but one – the causal factor of interest. In this situation, the "control" is the case without the intervention.

Experimental research designs usually achieve variation through time and across space, thus maximizing leverage into the fundamental problem of causal inference. They also minimize ceteris paribus assumptions, inherent in all causal analysis. First, because the intervention is manipulated by the researcher it is unlikely to be correlated with other things that might influence the outcome of interest. Thus, any changes in Y may be interpreted as the product of X and only X, other factors being held constant. (Natural interventions are likely to be accompanied by other factors that violate the ceteris paribus assumption.) Second, treatment and control cases are identical in all respects (except the intervention itself) that might affect the causal inference in question. This is usually achieved by a randomization of treatment and control groups. (However, randomization is not viewed here as a definitional attribute of experimental research designs.) Finally, the treatment and control groups are *isolated* from each other, preventing spatial contamination. This, again, means that the ceteris paribus assumption inherent in all causal inference is relatively safe. The control may be understood as reflecting a vision of reality as it would have been without the specified intervention.

Let us now reconstruct the logic of within-case research design through the logic of the classic experiment, which I shall define – stipulatively – as characterized by a manipulated intervention (the treatment) and a suitably matched control group. This suggests three parameters upon which all research designs may be evaluated: whether there is change in the status of the key causal variable during the period under observation (an intervention); whether this intervention is manipulated or not (i.e., whether the study is experimental or observational); and whether there is a well-matched control group. The intersection of these three dimensions produces a six-fold typology (not all logically conceivable cells are relevant), which I shall label as follows: 1) the Classic Experiment, 2) the Experimental Intervention, 3) the Natural Experiment, 4) the Natural Intervention, 5) the Natural Comparison, and 6) the Counterfactual, as described in Table 1. (Note that these terms carry a more specific meaning than they do in ordinary social science discourse; hence, the importance of capitalization.)

In order to familiarize ourselves with the differences among these six paradigmatic research designs I begin with a series of scenarios built around a central (hypothetical) research question: Does the change from a first-past-the-post (FPP) electoral system to a list-proportional (list-PR) electoral system moderate inter-ethnic hostility in a polity with high levels of ethnic conflict? I shall assume that one can effectively measure inter-ethnic hostility through a series of polls administered to a random sample (or panel) of respondents at regular intervals throughout the research period. This measures the outcome of our hypothetical study, the propensity to ethnic conflict. With this set-up, how might one apply the six foregoing designs?

In its simplest form, the Classic Experiment (#1) would proceed by the selection of two communities that are similar in all respects including the employment of a majoritarian electoral system and relatively high levels of inter-ethnic hostility. The researcher would then administer an electoral system change in one of these communities, holding the other constant. The final step would be to compare the results to see if there is a difference over time between treatment and control groups.

An Experimental Intervention (#2) would follow the same procedure, but without the control group. Consequently, the researcher's judgment of results would rest solely on a before/after comparison of inter-ethnic conflict in the community that underwent change in their electoral system.

A Natural Experiment (#3) is identical to the Classic Experiment except that the researcher is now operating in a non-experimental setting. This means that she must find two communities that are similar in all respects including the employment of a majoritarian electoral system and relatively high levels of inter-ethnic hostility, one of which changes its electoral system from majoritarian to proportional. She may then compare results across the two communities.

The Natural Intervention (#4) replicates the conditions of the second research design but in a non-experimental setting. That is, the researcher observes a community with a majoritarian electoral system and high levels of inter-ethnic hostility that undergoes an electoral system change to PR, comparing results before and after the intervention.

The Natural Comparison (#5) is identical to the third research design except that in this instance there is no intervention. Here, the researcher searches for two communities similar in all respects including the employment of a majoritarian electoral system and relatively high levels of inter-ethnic hostility. One employs a majoritarian electoral system and the other a proportional electoral system. This spatial variation on the key variable forms the crux of causal inference, but is not observable through time.

In a Counterfactual research design (#6), finally, the researcher observes a community with a majoritarian electoral system and high levels of inter-ethnic hostility that does *not* undergo an electoral system change to PR. Since there is no observable change over time in the key variable of interest,

her only leverage on this question is the counterfactual: what would have happened if this country had reformed its electoral system?

The essential properties of these six research designs are illustrated in Table 1, where Y refers to the outcome of concern, X_1 marks the independent variable of interest, and X_2 represents a vector of controls (other relevant exogenous factors that might influence the relationship between X_1 and Y). These controls may be directly measured or simply assumed (as they often are in randomized experiments). The initial value of X_1 is denoted "-" and a change of status as "+." The vector of controls, by definition, remains constant. A question mark indicates that the value of the dependent variable is the major objective of the analysis. Observations are taken before (t_1) and after (t_2) an intervention and are thus equivalent to pre- and post- tests.

Interventions may be manipulated (experimental) or natural (observational), as noted in Table 1. Note also that the nature of an intervention may be sudden or slow, dramatic or miniscule, dichotomous or continuous, and the effects of that intervention may be immediate or lagged. For ease of discussion, I shall assume that the intervention is of a dichotomous nature (present/absent, high/low, on/off), but the reader should keep in mind that the actual research situation may be more variegated. Thus, I use the term intervention (aka "event" or "stimulus") in the broadest possible sense, indicating any sort of change in trend in the key independent variable, X_1. It should be underlined that the absence of an intervention does not mean that a case does not change over time; it means simply that it does not experience a change of *trend*. Any evaluation of an intervention involves an estimate of the baseline – what value a case would have had without the intervention. A "+" therefore indicates a change in this baseline trend.

Because interventions may be multiple or continuous within a single case it follows that the number of temporal observations within a given case may also be extended indefinitely. This might involve a very long period of time (e.g., centuries) or multiple observations taken over a short period of time. Observations are thus understood to occur temporally within each case (t_1, t_2, t_3, . . . t_n).

Although the number of cases in the following examples varies, and is sometimes limited to one or two, research designs may incorporate any number of cases. In the previous example, each respondent to the survey of inter-ethnic conflict is understood as a case; there is evidently no limit, a priori, to the number of respondents that might be polled. Thus, the designations "treatment" and "control" in Table 1 may be understood to refer to individual cases *or* to groups of cases. (In this paper, the terms "case" and "group" will be used interchangeably.)[50]

Finally, the classical division of an experiment into two groups – a treatment and control – may be varied. There may, indeed, be a much larger number of groups, each receiving a slightly different treatment. At the limit, the treatments may be so variegated, and so numerous, as to defy a simple division into groups. Here, the researcher may choose to model the treatment in a general format -- usually a standard mathematical algorithm, which may be linear or non-linear. In this fashion, experiments merge with statistics. (Note, once again, the softness of the boundaries.)

In numbering these research designs (#1-6) I intend to indicate a gradual "falling away" from the experimental ideal. However, it would be incorrect to assume that a higher number necessarily indicates an inferior research design. In particular, it should be underlined that my discussion focuses mostly on issues of internal validity; often, the search for greater external validity leads to the adoption of an observational research design. Evidently, the three dimensions that define this typology do not exhaust the features of a good research design (Gerring 2001; 2006: chs 4-5). However, in most social-science

50 One obvious drawback to a very small sample is that one cannot randomize the selection of treatment and control cases. If there is only one control case, or several, it makes no sense to select it randomly (Gerring 2006: ch 5). Here, the case-selection procedure should follow the *most-similar* design.

research settings, and with a strong ceteris paribus caveat – i.e., when the chosen cases are equally representative (of some population), when the interventions are the same, and when other factors that might affect the results are held constant – the researcher will usually find that this numbering system accurately reflects the preferred research design. The six-part typology is intended to simplify the field of choices, expose the full range of options, and clarify the methodological issues attached to each one. Although initially the presentation may seem a trifle abstract it is hoped that after rehearsing numerous examples these models will begin to seem second-nature. It is also hoped that they will help the reader to craft her research and explain her choices in the simplest and clearest fashion possible. To reiterate, the essential questions are a) how experimental is your research design and b) in what specific ways does it deviate from the experimental ideal?[51]

51 There is one missing ingredient in this six-part typology. It concerns situations in which relevant observations are not comparable to one another and hence cannot be arrayed in a typical (large-N or small-N) research design. These sorts of observations have been referred to as causal-process (Brady 2004) or process-tracing (Gerring and Thomas 2005) observations.

Table 1. A Comprehensive Typology of Research Designs

Hypothesis: A change from FPP to list-PR mitigates ethnic hostility.

			t_1	t_2	
EXPERIMENTAL . . .					
1. Classic Experiment	Treatment	Y	--	?	Two similar communities with FPP electoral systems and high ethnic hostility, one of which is induced to change from FPP to list-PR. Ethnic hostility is compared in both communities before and after the intervention.
		X_1	--	+	
		X_2	--	--	
	Control	Y	--	?	
		X_1	--	--	
		X_2	--	--	
			t_1	t_2	
2. Experimental Intervention	Treatment	Y	--	?	A community with a FPP electoral system and high ethnic hostility is induced to change from FPP to list-PR. Ethnic hostility is compared before and after the intervention (identical to #1 except there is no control case).
		X_1	--	+	
		X_2	--	--	
OBSERVATIONAL . . .					
			t_1	t_2	
3. Natural Experiment	Treatment	Y	--	?	Two similar communities with FPP electoral systems and high ethnic hostility, one of which changes from FPP to list-PR. Ethnic hostility is compared in both communities before and after the intervention (identical to #1 except that treatment is not manipulated).
		X_1	--	+	
		X_2	--	--	
	Control	Y	--	?	
		X_1	--	--	
		X_2	--	--	
			t_1	t_2	
4. Natural Intervention	Treatment	Y	--	?	A community with a FPP electoral system and high ethnic hostility changes to list-PR. Ethnic hostility is compared before and after the intervention (identical to #2 except the intervention is not manipulated).
		X_1	--	+	
		X_2	--	--	
			t_1	t_2	
5. Natural Comparison	Treatment	Y		?	Two similar communities, one of which has PFF and the other list-PR. Ethnic hostility is compared in both communities (identical to #3 except there is no observable intervention).
		X_1		+	
		X_2		--	
	Control	Y		?	
		X_1		--	
		X_2		--	
			t_1	t_2	
6. Counterfactual	Control	Y		?	A community with a FPP electoral system and high ethnic hostility is considered, by counterfactual thought-experiment, to undergo a change to list- PR (identical to #4 except there is no treatment case).
		X_1		--	
		X_2		--	

Cases:
Treatment = with intervention
Control = without intervention

Variables:
Y = outcome
X_1 = independent variable of interest
X_2 = a vector of controls

Observations:
t_1 = pre-test (before intervention)
t_2 = post-test (after intervention)

Cells:
| = intervention
- = stasis (no change in status of variable)
+ = change (variable changes value or trend alters)
? = the main empirical finding: Y changes (+) or does not (-)

Three Genres of Causal Analysis

I agree with naturalists such as King, Keohane, and Verba that there is – or at least ought to be – one logic of inference that unites qualitative and quantitative work. I do not want to see the development of a separate and independent "qualitative methodology," in other words. Nor do I believe that this is possible or likely so long as we retain sight of the scientific ideal. If knowledge is to be systematic, parsimonious, cumulative, and replicable, if it is to extend to causal as well as descriptive inference, and if it is to strive for generality – if all of these scientific goals are to be respected then it makes no sense to develop separate fiefdoms for qualitative and quantitative methods. Both should speak to one another. And in order to facilitative this cross-field communication we need a common logic of inference.

That said, I also agree with the critics of DSI and other naturalistically-inclined methodologists: the current mainstream view of methods is often too narrow, too constraining, defining out much of what is now regarded as sound (and scientific) practice on the qualitative side of the ledger. This oversight is not, I think, malicious. My impression is that quantitative methodologists simply do not understand what constitutes a non-mathematical approach to empirical knowledge. Nor, for that matter, do most scholars who perform qualitative work. They conduct research on an intuitive level, but without the self-conscious tools of a "methodology." Indeed, they are often openly contemptuous of any attempt to intellectualize and systematize the work of scholarship. So it is a misunderstanding that – appropriately, in view of my thesis – crosses the qualitative/quantitative boundary.

What, then, *is* the qualitative/quantitative distinction? I would argue that it is best understood as derivative of an underlying methodological issue that remains obscured in most discussions. In my view, it is all about data *comparability*. Quantitative work presumes a high level of comparability among observations (pieces of evidence); qualitative work presumes a low level of comparability. This is the principal methodological justification for doing work that is quantitative or qualitative.

Accordingly, the methodological issues faced by research designs employed in causal analysis are recognizable by the number of comparable observations that lie within each "sample." Three broad categories are distinguishable: large-N samples, small-N samples, and samples of 1. This provides the empirical foundation and methodological rationale for three well-established styles of empirical research: 1) *Mathematical*, 2) *Comparative*, and 3) *Process-tracing*.

Table 2 illustrates the defining features of these genres, most of which follow, more or less ineluctably, from differences in sample size. Since these are extraordinarily broad groupings, encompassing all disciplines in the social sciences, and since the categories themselves are internally diverse, it seems appropriate to refer to them as methodological *genres*. In any case, it should be clear that when speaking about "Mathematical methods" or "Comparative methods" we are speaking about a diverse set of approaches.[52]

52 It should be clarified, finally, that this tripartite typology refers to methods of *data analysis*, not to methods of case selection or data generation. Prior to data analysis, we assume that researchers have carefully selected cases (either randomly or purposefully), and that researchers have generated data appropriately (either by experimental manipulation or some natural process). This data may contain quasi-experimental characteristics or it may be far from the experimental ideal. Data analysis may be conducted across cases or within cases. For our purposes, these issues are extraneous, though by no means trivial. In by-passing them I do not intend to downplay them. My intention, rather, is to focus narrowly on what analysts do with data once cases have been chosen, the data has been generated, and the relevant observations have been defined. This topic, I believe, is much less well understood.

Table 2. Three Genres of Causal Analysis

	Mathematical	**Comparative**	**Process Tracing**
Individual obs:	Quantitative	Quant or Qual	Quant or Qual
Groups of obs:	Large-N sample (comparable)	Small-N sample (comparable)	Disparate N=1 observations (non-comparable)
Total number of obs:	Large	Small	Indeterminate
Presentation of obs:	Rectangular dataset	Table or prose	Prose
Analytic technique:	Statistics, Boolean algebra	Most-similar, Most-different	Processual, Counterfactual, Pattern-matching Highly deductive
Covariation:	Real	Real	Real and imagined
Stability, replicability:	High	Moderate	Low
Familiar labels:	Statistics, QCA	Comparative, Comparative-historical, Small-N cross-case study	Historical, Narrative, Ethnographic, Legal, Journalistic, Single-case study

Mathematical Methods. The Mathematical genre will be familiar to most readers because it is represented by hundreds of methods textbooks and courses. Here, the analysis is typically conducted upon a large sample of highly comparable observations contained in a standard rectangular dataset, using some mathematical algorithm to establish covariational patterns within the sample. For better or worse, this is the standard template upon which contemporary understandings of research design in the social sciences is based. For some, it appears to be the *sine qua non* of social science research (Beck 2004; Blalock 1982, 1989; Goldthorpe 2000; King, Keohane, Verba 1994; Lieberson 1985; for general discussion see Brady and Collier 2004).

Our use of the term "Mathematical" does not presuppose any particular assumptions about how this analysis is carried out. If statistical, the model may be linear or non-linear, additive or non-additive, static or dynamic, probabilistic or deterministic (i.e., employing necessary causal factors), and so forth. The only assumption that statistical models must make is that the observations are *comparable* to one another – or, if they are not, that non-comparabilities can be corrected for by the modeling procedure (e.g., by weighting techniques, selection procedures, matching cases, and so forth). For statisticians, the assumption of unit homogeneity is paramount. It should be clear that the same requirements apply whether the observations are defined spatially (a cross-sectional research design), temporally (a time-series research design), or both (a time-series cross-section research design). By extension, the same requirements apply whether the analysis is probabilistic ("statistics") or deterministic (as in some versions of Qualitative Comparative Analysis [Ragin 1987, 2000]).

As a rule, Mathematical work employs a sample that remains fairly stable throughout the course of a single study. Granted, researchers may exclude or down-weight outliers and high-leverage observations, and they may conduct sub-sample analyses. They may even interrogate different datasets in the course of a longer study, or recode the sample to conduct sensitivity analyses. However, in all these situations there is a relatively explicit and well-defined sample that contains the evidentiary basis for causal inference. The importance of this issue will become apparent as we proceed.

Comparative Methods. The two most familiar Comparative methods are *most-similar* analysis (a.k.a method of agreement) analysis and *most-different* analysis (aka method of difference), both of which can be traced back to J.S. Mill's nineteenth-century classic, *System of Logic* (1834/1872). In most-similar analysis, cases are chosen so as to be similar on all irrelevant dimensions and dissimilar on both the hypothesized causal factor and the outcome of interest. In most-different analysis, cases are chosen to maximize difference among the cases on all causal factors (except one), while maintaining similarity on the outcome. The most-similar research design is more common, and probably more well-grounded, than the most-different research design (Gerring 2006: ch 5; Seawright and Gerring 2005).

The details of these research designs are not important. What is important is that the cross-case component of the analysis be fairly explicit. There must be a recognizable sample within which the chosen cases are analyzed. In other words, there must be significant cross-case variation and this variation must comprise an important element of the overall analysis. This is the "comparative method" as it has become known within the subfield of comparative politics (Collier 1993).[53] "Comparative-historical" work is similar to the foregoing except that the analysis also incorporates a significant over-time component (Mahoney and Rueschemeyer 2003).

53 My use of the term Comparative includes what Mahoney (1999) labels "nominal comparison" and "ordinal comparison," but not what he labels "narrative analysis," which I incorporate under Process Tracing below.

Cases are thus examined spatially and temporally, and the temporal analysis usually includes a change in one or more of the key variables, thus introducing an intervention ("treatment") into the analysis.[54]

Comparative methods, like Mathematical methods, are based upon a relatively stable sample of comparable cases. Granted, there are likely to be some shifts in focus over the course of a longer study. Sometimes, a researcher will choose to focus on a series of nested sub-samples, e.g., paired comparisons (Collier and Collier 1991). The small size of the sample means that any change in the chosen cases will have a substantial impact on the sample, and perhaps on the findings of the study. Ceteris paribus, small samples are less stable than large samples.

Because Comparative methods must employ cases that are fairly comparable to one another, they may be represented in a standard, rectangular dataset where the various dimensions of each case are represented by discrete variables. Yet, because there are relatively few cases (by definition), it is rare to see a dataset presentation of the evidence. Instead, scholars typically rely on small tables, 2x2 matrices, simple diagrams, or prose.

The most important difference between Mathematical methods and Comparative methods is that the latter employs small samples that may be analyzed without the assistance of interval scales and formal mathematical models. This does not *preclude* the use of mathematical models (e.g., Houser and Freeman 2001), or of algorithms to assign precise weightings to "synthetic" cases (Abadie and Gardeazabal 2003). However, it is not the usual mode of procedure. Indeed, statistics are relatively powerless when faced with samples of a dozen or less. A simple bivariate analysis may be conducted, but this does not go much further than what could be observed visually in a table or a scatterplot diagram.

Another difference with the Mathematical framework is that Comparative methods presuppose a fairly simple coding of variables, usually in a dichotomous manner. Similarities and differences across cases must be clear and distinct, otherwise they cannot be interpreted (due to the small-N problem). Thus, continuous variables are usually dichotomized into high/low, present/absent, strong/weak, and so forth. Simple coding schemes, and the absence of probability distributions, impose a deterministic logic on Comparative methods, such that causal factors (or combinations of factors) must be understood as necessary, sufficient, or necessary and sufficient. Deterministic assumptions may also be employed in Mathematical methods, particularly Boolean methods, but they are not de rigueur in statistical methods. Moreover, the smaller the sample size, the more difficult it is to incorporate continuous causal factors and probabilistic logic if firm conclusions are to be reached.

Process Tracing Methods. Process Tracing, in our lexicon, refers to any method in which the researcher analyzes a series of noncomparable observations occurring within a single case.[55] Studies that employ Process Tracing typically consist of many observations (either qualitative or quantitative), each making a slightly different point, but all related to some overall argument (i.e., the primary inference).

54 My discussion thus far has approached Comparative methods according to the primary unit of analysis, usually referred to as a "case" (a spatially and temporally delimited unit that lies at the same level of analysis as the principal inference). To be sure, this genre of work may also exploit *within-case* variation, which might be large-N (e.g., a mass survey of individual respondents or a time-series analysis of some process), small-N (e.g., a comparison among a half-dozen regional units), or a series of N=1 observations (e.g., a study of a particular decision or set of decisions within the executive). In short, the within-case components of Comparative methods are indeterminate; they may be Mathematical, Comparative, or Process Tracing. The fact that a single study may employ more than one method is not disturbing; as we observed, a change in an author's level of analysis *often* corresponds to a change in research design. In short, the same tripartite typology is applicable at any single level of analysis; but it does not always apply across-the-board to all levels of analysis in a given study.

55 The term "process tracing" is ambiguous, having been appropriated for a variety of uses. For some writers, it refers to any investigation (qualitative or quantitative) into causal mechanisms (George and Bennett 2005). There is, to be sure, a strong affinity between this technique, as we describe it, and a researcher's insight into causal paths. However, it may be a mistake to *define* process tracing as the search for causal mechanisms. After all, this is also an objective of Mathematical and Comparative studies. In short, while Process-Tracing methods give more attention to causal mechanisms, this should not be considered a defining feature. Our definition of process tracing might also be labeled "causal-process" observations (following Brady and Collier 2004), or alternatively, colligation, narrative explanation, pattern-matching, sequential explanation, genetic explanation, and causal chain explanation. For general discussion, see Brady (2004), George and Bennett (2005: ch 8), Little (1995: 43-4), Scriven (1976), Seawright and Collier (2004), Tarrow (1995: 472). For examples, see Geddes (2003: ch 2), George and Smoke (1974), Goldstone (2003: 50-1), George and Bennett (2005: appendix).

Since the observations are not comparable with one another, the presentation is delivered in prose – or what Mahoney (1999) labels "narrative analysis." However, it is the absence of comparability among adjacent observations – not the use of prose (or narrative) – that makes this approach so distinctive, and so mysterious. Process-Tracing methods do not conform to standard notions of methodological rigor because most elements of a "research design," in the usual sense of the term, are absent. There is, for example, no formally defined sample of observations, as with Mathematical and Comparative methods. Moreover, the methods for making causal inferences that link observations into a causal chain are often not explicitly stated. Consequently, Process-Tracing studies give the impression of being informal, ad hoc -- one damn observation after another.

The skepticism of mainstream methodologists is not difficult to comprehend. William Riker (1985: 62-3; see also Beck 2004) regards process-tracing as "scientifically impossible." Tracing a process, and imposing a pattern is, of course, no more and no less than writing history. Although some nineteenth-century historians claimed to be scientific, such a claim has seldom been put forward in this century until now, when it rises up, camouflaged, in social science. There was good reason for abandoning the claim: Historical explanation is genetic. It interprets cause as no more than temporal sequence, which, in the philosophy of science, is precisely what has long been denounced as inadequate. Causality in science is a necessary and sufficient condition; and, although temporal sequence is one of several necessary conditions, it is not sufficient. . . Process-tracing of the history of an event, even the comparison of several traced processes, does not give one generalizations or theory. However, we shall argue that the wayward reputation of Process Tracing is only partially deserved. Indeed, inferences drawn from Process-Tracing methods may be more secure, at least in some instances, than inferences based on Mathematical or Comparative methods. Thus, there are strong arguments for the employment of non-comparable (N=1) observations in social science.

We begin with an extended example drawn from Henry Brady's (2004: 269-70) reflections on his study (in tandem with a team of methodologists) of the Florida election results in the 2000 presidential election. In the wake of this close election at least one commentator suggested that because several networks called the state for Gore prior to a closing of the polls in the Panhandle section of the state, this might have discouraged Republican voters from going to the polls, and therefore might have affected the margin (which was razor thin and bitterly contested in the several months after the election) (Lott ??). In order to address the question, Brady stitches together isolated pieces of evidence in an inferential chain. He begins with the timing of the media calls – ten minutes before the closing of the polls in the Panhandle. "If we assume that voters go to the polls at an even rate throughout the day," Brady continues, "then only 1/72nd (ten minutes over twelve hours) of the [379,000 eligible voters in the panhandle] had not yet voted when the media call was made." This is probably a reasonable assumption. ("Interviews with Florida election officials and a review of media reports suggest that, typically, no rush to the polls occurs at the end of the day in the panhandle.") This means that "only 4,200 people could have been swayed by the media call of the election, if they heard it." He then proceeds to estimate how many of these 4,200 might have heard the media calls, how many of these who heard it were inclined to vote for Bush, and how any of these might have been swayed, by the announcement, to go to the polls in the closing minutes of the day. Brady concludes: "the approximate upper bound for Bush's vote loss was 224 and . . . the actual vote loss was probably closer to somewhere between 28 and 56 votes."

Brady's conclusions rest not on a formal research design but rather on isolated observations combined with deductive inferences: How many voters "had not yet voted when the media called the election for Gore? How many of these voters heard the call? Of these, how many decided not to vote? And of those who decided not to vote, how many would have voted for Bush?" (Brady 2004: 269).

This is the sort of detective work that fuels the typical Process-Tracing study, and it is not a sort that can be represented in a rectangular dataset. The reason is that the myriad pieces of evidence are not comparable to each other. They all support the central argument – they are not "random" – but they do not comprise observations in a larger sample. They are more correctly understood as a series of N=1 (one-shot) observations – or perhaps the more ambiguous phrase "pieces of evidence" is appropriate. In any case, Brady's observation about the timing of the call – ten minutes before the closing of the poll – is followed by a second piece of evidence, the total number of people who voted on that day, and a third and a fourth. It would be impossible to string these together into a large, or even moderately-sized, sample, because each element is disparate. Being disparate, they cannot be counted. While the analytic procedure seems messy, we are convinced by its conclusions – more convinced, indeed, than by the large-N analysis that Brady is arguing against (in which . . .). Thus, it seems reasonable to suppose that, in some circumstances at least, Process Tracing is more scientific than sample-based inferences, even though its method is difficult to describe.

This is the conundrum of Process-Tracing research. We are often convinced by the results, but we cannot explain – at least not in any generalizable, formal fashion – why. Our confidence appears to rest on highly specific propositions and highly specific observations. There is little we can say, in general, about "Brady's research design" or other Process-Tracing research designs. It is no surprise that Process Tracing receives little or no attention from traditional methods texts, structured as they are around the quantitative template (e.g., King, Keohane, and Verba 1994). These methods texts do not tell us why a great deal of research in the social sciences, including a good deal of case study research, succeeds or fails.

While sample-based methods (both Comparative and Mathematical) can be understood according to their covariational properties, Process-Tracing methods invoke a more complex logic, one that is analogous to detective work, legal briefs, journalism, traditional historical accounts, and single-case studies. The analyst seeks to make sense of a congeries of disparate evidence, some of which may explain a single event or decision. The research question is always singular, though the ramifications of the answer may be generalizable. Who shot JFK? Why did the US invade Iraq? What caused the outbreak of World War One? Process-Tracing methods are, by definition, case-based. If a researcher begins to draw comparisons with other assassinations or other wars, then she is using (at least implicitly) a Comparative method, which means that all the standards of rigor for Comparative methods pertain and the researcher is entering a different methodological context.

It is important to note that the observations enlisted in a Process-Tracing case study may be either qualitative or quantitative. Brady employs a good deal of quantitative evidence. However, because each quantitative observation is quite different from the others they do not collectively constitute a sample. Each observation is sampled from a different population. This means that each quantitative observation is qualitatively different. Again, it is the comparability of adjacent observations, and the number of those observations, not the nature of the observations, that define a study as Mathematical, Comparative, or Process Tracing.

Note also that because each observation is qualitatively different from the next, the entire set of observations in a Process-Tracing study is indeterminate and unstable. The "sample" (we use this term advisedly) shifts from observation to observation. Because of this, we refer to samples of 1, or N=1 observations (of which there may be many in a single case study). A careful reader might object that the notion of an "observation" implies the existence of other comparable observations in a larger population. We accept that this is true for most observations. The issue is not whether comparable observations exist, but rather whether those other observations are considered (i.e., sampled and analyzed) in the case study. If they are not considered, then we have a set of N=1 observations. Regardless of how carefully

one seeks to define these things, there should be no disagreement on our basic point that samples, populations, and sampling techniques are not well specified in Process-Tracing methods. If they are well specified, then we are working in the realm of Comparative or Mathematical methods.

There may be *many* non-comparable observations in a single Process-Tracing study, so the cumulative number of observations could be quite large. However, because these observations are not well defined, it is difficult to say exactly how many there are. Non-comparable observations are, by definition, difficult to count. Recall, from our previous discussion, that the act of counting presumes comparability among the things being counted. Process-Tracing evidence lacks this quality; this is why it is resistant to the N question. In an effort to count, one may of course resort to lists of what appear to be distinct pieces of evidence. This approximates the numbering systems commonly employed in legal briefs. But lists can always be composed in multiple ways, so the total number of observations remains an open question. We do not know, and by the nature of the analysis cannot know, precisely how many observations are present in studies such as Fenno's *Homestyle* (1978), Kaufman's *The Forest Ranger* (1960), Geertz's *Negara* (1980), and Pressman and Wildavsky's *Implementation* (1973). Process-Tracing observations are not different examples of the same thing; they are, instead, *different things*. Consequently, it is not clear where one observation ends and another begins. They flow seamlessly together. Thus, we cannot re-read Fenno, Kaufman, Geertz, or Pressman and Wildavsky with the aid of a calculator and hope to discover their true N, nor would we gain much – if any – analytic leverage by so doing. Quantitative researchers are inclined to assume that if observations cannot be counted they must not be there, or – more charitably – that there must be very few of them. Qualitative researchers may insist that they have many "rich" observations at their disposal, which provide them with the opportunity for "thick" description; but they are unable to say, precisely, how many observations they have, or where these observations are, or how many observations are needed for thick analysis. Indeed, the observations themselves remain undefined.

This ambiguity is not in our opinion troublesome, for the number of observations in a Process-Tracing study does not bear directly on the usefulness or truthfulness of that study. While the number of observations in a sample drawn from a well-defined population contains information directly relevant to any inferences that might be drawn from that sample, the number of observations in a Process-Tracing study (assuming one could estimate their number) has no obvious relevance to inferences that might be drawn from that study. Consider that if it was merely quantity that mattered we might safely conclude that longer studies, which presumably contain more observations, are more reliable or valid than shorter studies. Yet, it is laughable to assert that long books are more convincing than short books. It is quite evidently the quality of the observations and how they are analyzed, not the quantity of observations, that is relevant in evaluating the truth-claims of a Process-Tracing study.

Thus, the N=1 designation that we have attached to Process-Tracing evidence should not be understood as pejorative. In some circumstances, one lonely observation (qualitative or quantitative) is sufficient to prove an inference. This is quite common, for example, when the author is attempting to reject a necessary or sufficient condition. If we are inquiring into the cause of Joe's demise, and we know that he was shot at close range, we can eliminate suspects who were not in the general vicinity. One observation – "I saw Peter at the supermarket" – is sufficient to provide fairly conclusive proof (provided, of course, that the witness is reliable). Better yet would be a videotape of the suspect at the supermarket from a surveillance camera. This would be conclusive evidence to falsify a hypothesis (in this case, Peter shot Joe), even though it is not quantitative or comparable evidence.

Process-Tracing methods apply only to situations in which the researcher is attempting to reconstruct a sequence of events occurring within a single case – i.e., a relatively bounded unit such

as a nation, family, legislature, or decision-making unit. That case may be quite broad, and might even encompass the whole world, but it must be understood as a single unit, for purposes of the analysis. All Process-Tracing methods are inherently within-case analysis. If several cases are analyzed, the researcher has either switched to a different style of analysis or adopted an additional style of analysis, one in which there is a specifiable sample (either large-N or small-N). The researcher may, for example, have begun with a Process-Tracing analysis within one case study, and later switched levels of analysis by comparing that case study with other case studies using a Comparative method.

What is it, then, that makes a Process-Tracing study convincing or unconvincing? What are the methods within this genre of causal analysis? A fundamentally puzzling aspect of the Process-Tracing method is that it rests, at times, on extremely proximate evidence (observations lying close to the "scene of the crime"), and at other times on extremely general assumptions about the theory at hand or the way the world works. Process Tracing thus lies at both extremes of the inductive-deductive spectrum. Sample-based studies, by contrast, generally require fewer deductive assumptions and, at the same time, are more removed from the facts of the case. The extreme quality of Process Tracing – which bounces back and forth from Big Theory to detailed observation – contributes to its "unstable" reputation. However, there are good reasons for this back-and-forth.

Broadly, we may distinguish among two styles of Process-Tracing research; one is *exploratory* and the other *confirmatory* (Gerring 2001: ch ?). In an exploratory mode, the researcher seeks to discover what went on in a specific context without any strong theoretical preconceptions. The question "What happened?" is asked in an opened-ended fashion. While this may seem removed from the deductive mode of inquiry that we have described, in fact it relies heavily on an understanding (theoretical or pre-theoretical) of the way the world works. In order to demonstrate a causal relationship from the mass of evidence at hand it is necessary to provide a reconstruction of the event under slightly different (imaginary) circumstances. One must construct valid "what if?" scenarios. The method of Process Tracing is thus linked to what has come to be known as the counterfactual thought-experiment (cites). There is simply no other way that the tracing of a single process through time can make causal claims – since, by definition, there are no "real" (actually existing) contrasting cases. Note that if there are other cases, and if these cases are brought into the analysis, then the researcher has transitioned into either a Mathematical or Comparative mode of analysis (depending upon the number of comparison-cases she is considering and her mode of examination). Process Tracing is limited to a single thread of occurrences. To be sure, the fact that these occurrences can be interpreted at all is courtesy of the analyst's general assumptions about how the world works (or how this particular part of the world works). This is why general knowledge – even if it is not specific to a particular theory – counts heavily in all Process-Tracing studies. The conjunction of general and specific knowledge is nicely brought out in Clayton Roberts's (1996: 66) description of process tracing as "the minute tracing of the explanatory narrative to the point where the events to be explained are microscopic and the covering laws correspondingly more certain." While we hesitate to invoke the rather controversial notion of a covering law, we hold, with Roberts, that Process Tracing conjoins highly specific and highly general observations.

Confirmatory Process Tracing also relies on imaginary counterfactuals, and also combines the general and the specific. The difference is that here a theory, rather than one's general knowledge of the world, is instrumental in identifying relevant factuals and counterfactuals. This style of Process Tracing sometimes goes under the label of "pattern-matching." Here, a theory "generates predictions or expectations on dozens of other aspects of the [subject at hand], and [the writer] does not retain the theory unless most of these are also confirmed. In some sense, he has tested the theory with degrees of freedom coming from the multiple implications of any one theory" (Campbell 1975/1988: 380; see

also Scriven 1976). An exploratory study asks "What happened?" A pattern-matching investigation inquires, first, "What should have happened if Theory X is true?" and, second, "Did that predicted course of action actually occur?" To be sure, in practice researchers often blend these two closely related techniques. A researcher may start inductively, but find herself with several weak links in the causal chain. To bolster these links, she might turn to pattern-matching, using hypotheses drawn from theories (i.e., covering laws) to make the causal inferences for those links.

References

NB: The foregoing discussion draws from the following works, as well as from work by other scholars (e.g., Andrew Bennett, Henry Brady, David Collier, Colin Elman, Jim Mahoney) and discussions with many friends and associates.

Abadie, Alberto and Javier Gardeazabal. (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* (March): 113-32.

Beck, Nathaniel. (2004). "Is Causal-Process Observation an Oxymoron? A Comment on Brady and Collier (eds.), 'Rethinking Social Inquiry.'" *Ms.*

Blalock, Hubert M., Jr. (1982). *Conceptualization and Measurement in the Social Sciences.* Beverly Hills: Sage.

Blalock, Hubert M, Jr. (1989). "The Real and Unrealized Contributions of Quantitative Sociology." *American Sociological Review,* 54: 447-60.

Brady, Henry E. (2004). "Data-Set Observations versus Causal-Process Observations: The 2000 U.S. Presidential Election." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards,* edited by H. E. Brady and D. Collier. Lanham: Rowman & Littlefield.

Brady, Henry E. and David Collier (eds). (2004). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham: Rowman & Littlefield.

Campbell, Donald T. (1975/1988). "'Degrees of Freedom' and the Case Study." In *Methodology and Epistemology for Social Science*, edited by E. Samuel Overman. Chicago: University of Chicago Press.

Collier, David. (1993). "The Comparative Method." In *Political Science: The State of the Discipline II*, edited by A.W. Finifter. Washington, DC: American Political Science Association.

Collier, Ruth B. and David Collier. (1991). *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton: Princeton University Press.

Fenno, Richard F., Jr. (1978). *Home Style: House Members in their Districts.* Boston: Little, Brown.

Geddes, Barbara. (2003). *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor: University of Michigan Press.

Geertz, Clifford. (1980). *Negara: The Theatre State in Nineteenth-Century Bali*. Princeton: Princeton University Press.

George, Alexander L. and Andrew Bennett. (2005). *Case Studies and Theory Development*. Cambridge: MIT Press.

George, Alexander L. and Richard Smoke. (1974). *Deterrence in American Foreign Policy: Theory and Practice*. New York: Columbia University Press.

Gerring, John. (2001). *Social Science Methodology: A Criterial Framework.* Cambridge: Cambridge University Press.

Gerring, John. (2004). "What is a Case Study and What is it Good For?" *American Political Science Review,* 98: 341-54.

Gerring, John. (2005). "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17:2 (April).

Gerring, John. (2006). *Case Study Research: Principles and Practices.* Cambridge: Cambridge University Press (forthcoming).

Gerring, John and Craig Thomas. (2005). "What is 'Qualitative' Evidence?: When Counting Doesn't Add Up." *In process.*

Gerring, John and Jason Seawright. (2005). "Selecting Cases in Case Study Research: A Menu of Options." *In process.*

Gerring, John and Paul A. Barresi. (2003). "Putting Ordinary Language to Work: A Min-Max Strategy of Concept Formation in the Social Sciences." *Journal of Theoretical Politics,* 15: 201-32.

Gerring, John and Rose McDermott. (2005). "Experiments and Observations: Towards a Unified Framework of Research Design." *In process*.

Goldstone, Jack A. (2003). "Comparative Historical Analysis and Knowledge Accumulation in the Study of Revolutions." In *Comparative Historical Analysis in the Social Sciences*, edited by J. Mahoney and D, Rueschemeyer. Cambridge: Cambridge University Press.

Goldthorpe, John H. (2000). *On Sociology: Numbers, Narratives, and the Integration of Research and Theory.* Oxford: Oxford University Press.

Houser, Daniel and John Freeman. (2001). "Economic Consequences of Political Approval Management in Comparative Perspective." *Ms.*.

Kaufman, Herbert. (1960). *The Forest Ranger: A Study in Administrative Behavior*. Baltimore: Johns Hopkins University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Lieberson, Stanley. (1985). *Making it Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Little, Daniel. (1995). "Causal Explanation in the Social Sciences." *Southern Journal of Philosophy,* 34: 31-56.

Mahoney, James. (1999). "Nominal, Ordinal, and Narrative Appraisal in Macro-Causal Analysis." *American Journal of Sociology,* 104: 1154-96.

Mahoney, James and Dietrich Rueschemeyer (eds). (2003). *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Mill, John Stuart. (1843/1872). *System of Logic, 8th ed*. London: Longmans, Green.

Pressman, Jeffrey L. and Aaron Wildavsky. (1973). *Implementation*. Berkeley: University of California Press.

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

Ragin, Charles C. (2000). *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Riker, William H. (1985). "Comments on 'Case Studies and Theories of Organizational Decision Making." *Advances in Information Processing in Organizations*, 2: 59-64.

Roberts, Clayton. (1996). *The Logic of Historical Explanation*. University Park: Pennsylvania State University Press.

Seawright, Jason and David Collier. (2004). "Glossary." Pp. 273-313 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited H. E. Brady and D. Collier. Lanham: Rowman & Littlefield.

Scriven, Michael. (1976). "Maximizing the Power of Causal Investigations: The Modus Operandi Method." Pp. 101-18 in *Evaluation Studies Review Annual*, edited by G.V. Glass. Beverly Hills: Sage.

Tarrow, Sidney. (1995). "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review,* 89: 471-74

Conducting Intensive Interviews and Elite Interviews

Jennifer Hochschild
Department of Government
Department of African and African-American Studies
Harvard University

This memo focuses on the logic of conducting intensive interviews and elite interviews. By the former, I mean long (at least 1 hour, usually more) discussions with people chosen at random or in some other way that does not refer to them as specific individuals. By the latter, I mean discussions with people who are chosen because of who they are or what position they occupy. That is, by "elite" I do not necessarily mean someone of high social, economic, or political standing; the term indicates a person who is chosen by name or position for a particular reason, rather than randomly or anonymously.

A central purpose of an intensive interview is conceptual mapping: How and why does the ordinary person on the street think and feel about a set of issues or ideas? What connections do people make; what plausible linkages do they not see? Where do they express ambivalence, incoherence, certainty, passion, subtlety, and why? In contrast, a central purpose of elite interviews is to acquire information and context that only that person can provide about some event or process: What did that person do and why? How does he or she explain and justify his/her own behavior? What does the person remember of how others behaved, and why? How does the person understand and explain the trajectory of the event or process? What succeeded or failed, from that person's vantage point?

Both forms of interview are invaluable for a large swath of research questions in social science. Intensive interviews can be supplements to, inspirations for, or correctives of virtually any public opinion survey. Elite interviews can play the same multiple roles for most research that traces the history or development of a phenomenon over the past half century, roughly speaking. Thus while each is, in one sense, just a particular form of systematic qualitative research, together they are likely to be vital elements of almost any research program that engages with recent intentional human behavior.

Standards for Rigor

Intensive Interviews: The trick with qualitative interviews is to know how much and what aspects of the standards for survey research are applicable to this research method – *and* which standards are inappropriate. Some elements of survey design are valuable. These include 1) the desirability in many cases of obtaining respondents randomly rather than through convenience or snowball samples; 2) the desirability in many cases of presenting the same relatively neutral persona to respondents so that they engage with the issues at hand rather than with the interviewer as a person (but see below for a caveat); 3) the need for informed consent; 4) the need to avoid questions that are biased, leading, or otherwise likely to distort the respondents' reported views; 5) the need for a systematic and replicable way of making sense of the data after the interviews are collected; 6) and the importance of making the evidence publicly available to other researchers.

However, many elements of survey research design are not appropriate for intensive interviews; if they are used, these elements will confuse or even undermine the value of interviews. The tip-off is a discussion of generalization (inevitably defensive from a qualitative interviewer), or sentences that seek to show that more of X type of interviewees had a given response than of Y type of interviewees. That is, treating interview subjects like a very small survey sample is a mistake – it will not convince surveyors, and it brings to the fore the disadvantages rather than the advantages of this type of research.

Instead, intensive interviews should focus on doing just what surveys cannot do, that is, finding out how people frame their views, why they hold those views, and how they make connections or demonstrate disjunctions among discrete opinions. Intensive interviews can do directly what statistical analysis seeks to do indirectly and at a distance – show what attitudes or values are "correlated," how strongly they are associated, and how and why people link or morselize particular views. For example, rather than controlling for race on the assumption that African Americans view aspects of American politics differently from whites, one can ask directly how, why, and how much race matters when a respondent expresses a view. One can also consider that issue indirectly, by examining the respondent's level of certainty, comfort, and ease of explanation, or unstated assumptions about what seems obviously true or false (or good or bad), wealth of anecdotes or supporting evidence, and justifications for a view.

Thus standards for rigor in intensive interviews should focus on the degree to which the respondent can be induced to express views, perhaps even to develop them in the course of the interview, and to examine carefully what lies behind his or her own comments. The interviewer should focus especially – albeit tactfully -- on apparent inconsistencies, disconnections, or ambivalences in order to see, for example, if what appears to be ideological incoherence is simply a distinctive way of ordering or clustering particular values that does not map onto a liberal-conservative dimension. The interviewer may need to change his or her persona in order to get the fullest set of responses; I have had respondents who would only talk with me if I were willing to argue back and engage in a genuine conversation rather than a one-way probing. Similarly, the interviewer may need to change the order of the topics under discussion, to change question wording, to spend a lot (or very little) time on one topic compared with another, to make questions more or less abstract for a given respondent, to show emotional responsiveness—all anathema to a survey researcher but all possibly necessary for the task of getting the respondent to think carefully, fully, and openly about the issues at hand.

Analysis of interviews also requires appropriate standards of rigor. Interviews should always be transcribed in full, in my opinion, including hesitations and emphases, so that one can have the full array of responses always accessible. The great temptation is to pick and choose strong quotations that make the points the interviewer wanted to have made to begin with, and to string together a set of ideas from the respondent that cohere in a particular way. This need not be dishonest or even intentional; intensive interviews contain a lot of what appears to be "noise," and inevitably a great deal of material must be discarded in order to develop a coherent, thematic narrative reporting the results. But it is essential that the researcher allow anomalies, apparent inconsistencies, less savory aspects of the responses, even incoherence itself, to be part of the analysis and report. As I argued in *What's Fair?* at one point, a totally baffling paragraph may reveal a great deal about how a person thinks about a complex problem; similarly, a respondent who suddenly announces, as one of mine did, that we should kill off all the people in the world who disagree with him can change one's sense of what it might mean to be an deeply committed humanitarian liberal.

In short, rigor in intensive interviews is not the same as that for surveys, and may in some cases require the opposite strategy or behavior. It also requires the researcher to pay at least as much attention to what he or she does not like or did not expect or does not understand in analyzing the interview transcripts as to what seems to make sense along lines that were predicted (or predictable) before the research began.

Elite Interviews: Rigor in elite interviews is more straightforward, and more closely analogous to traditional journalists' ethics and rules of engagement. The interviewer must know as much as possible about the context, stance, and past behavior of the interview subject before beginning the conversation; that seems obvious for a member of Parliament or corporation president, but is equally true for a community organizer or foreman on the assembly line. One does not want to waste the respondent's

time, and one wants to get as complete, honest, and nuanced a story as possible from the respondent. Being able to say, "So, how does that accord with what you said X days ago or what you did Y years ago?" gives the interviewer credibility, and helps to keep the respondent from telling partial or –shall we say – imaginative narratives. It also enables the interviewer to probe more deeply into the respondent's perhaps idiosyncratic or nonrational stances, and gives the respondent more material with which to effectively develop his or her own explanation of past behavior.

The interviewer can *carefully* triangulate among respondents; without revealing any confidences or names of previous subjects, one can sometimes use information gleaned from a previous interview to question or push a current subject a little more deeply. The interviewer should also always ask the opposite question from the one just asked ("What was your most effective strategy to accomplish X?" and then, "What did you try in order to accomplish X that did not work as you intended?"). Finally, the same rules apply for interpreting these interviews as in qualitative interviews, or in conventional journalism: one must portray respondents fairly, give the reader enough evidence to show the complexities and problems in one's interpretation as well as its strengths, illuminate rather than distort the historical record as revealed by the respondents, and provide a plausible interpretation to pull all the threads together.

A final thought about interviews, especially elite interviews: it is tempting, particularly with well-educated or highly knowledgeable subjects, to ask them one's own research question. Even if the question is couched in layperson's language with a minimum of verbal flourishes, this is usually a mistake in my experience. Few interview subjects think in the ways that social scientists think, so posing one's own analytic puzzle to the subject usually just elicits puzzled stares and silence or stammers. More seriously, one purpose of this sort of interview is to leave enough space between the researcher's initial preconceptions or frameworks and the subjects' particular framework and vantage point so that the researcher stays open to surprise and anomaly. Doing too much to set up the interview in terms of one's own theoretical logic once again moves interviews too far in the direction of survey research.

Communicating Standards to Other Disciplines

A simple starting point would be more articles or book chapters laying out the logic of intensive and elite interviews. Such a document should focus on their distinctive qualities and include, among other things, an explicit discussion of how they are *not* like survey research, except with a long interview schedule and small N. It would similarly be helpful to distinguish these types of interviewing from ethnographic research, which seldom asks research subjects for self-conscious statements of values and attitudes in an artificial context. Teaching the same points in courses on social science methods would help also, of course.

The article or book chapter should also address the vexed question of how to interpret the results once interviews are complete. This is easier with elite interviews, since one is basically developing a history or analytic narrative; at least there is a chronological logic available as an initial analytic template. Intensive interviews are harder to interpret, and it is much harder to convey to someone else how to do it. Software is available for their analysis; in my view even the most sophisticated qualitative software provides a useful starting point but never suffices. In my experience, the process of developing themes and arguments out of transcripts of intensive interviews is endlessly iterative. The final argument emerges out of some combination of initial framework, unpredictable insight, multiple readings, engagement with the extant literature on the subject, and many draft pages. An article or

chapter that provides a chunk of unedited transcript, then walks the reader through the process of honing that material into finished prose would be very valuable (and hard to do!).

Topics Particularly Suited to Intensive and Elite Interviews

Intensive interviews have three broad purposes. First, they can provide the research material itself: topics such as group identity, individual ideology, attitudes about newly developing policy issues, explanations for political activism or social engagement, recounting of traumatic experience, and explications of relationships or emotions are all amenable to intensive interviews. Second, this sort of interview can be very useful for designing a theoretically elegant and empirically appropriate survey instrument. Similarly, it can be used to provide a context or set of insights to help a researcher make sense of results from surveys that have already been conducted. Third and most generally, intensive interviews are a vehicle for developing explanations for inevitably superficial survey results. That is, perhaps the survey is the pretest, conducted mainly to suggest areas of discussion for the intensive interviews to follow. In that logic, the qualitative interviews will confirm, disconfirm, or transform one's hypotheses; the surveys are mainly the set-up.

Elite interviews can have the same three purposes. As the research content itself, a set of these interviews is clearly appropriate for the study of recent historical change, process-tracing studies of policy enactment or implementation, the role of memory and perception in political or social activity, and the role of elites (broadly defined) in a political, social, or economic process. Second, elite interviews can function as a sort of pre-test to help one discern which institutions or processes should be carefully studied through some other means such as content analysis, formal modeling, or statistical manipulation. Third and most generally, elite interviews can give substance and meaning to prior analyses of institutions, structures, rule-making, or procedural controls. Knowing how an open or closed rule works in a Congressional committee, for example, is an essential starting point; talking with people whose political strategy depends on whether there is an open or closed rule gives depth to the more formal logic of Congressional decision-making. That is, elite interviews can play the same role with regard to institutional analysis that intensive interviews can play with regard to survey research: they can set up the alternative research strategies, or they can make sense of what has been gleaned from those strategies.

Thoughts on Defining and Communicating Standards of Rigor in Political Science

Kathleen Thelen
Department of Political Science
Northwestern University

This memo provides some thoughts (from the perspective of someone engaged in comparative historical-institutional research in political science) on the general issues we were asked to address, namely:[56]

1. What are the standards of rigor in your discipline?
2. How might these standards of rigor be communicated to or applied in other disciplines?
3. What areas or topics are most promising for investigation using qualitative methods?

The executive summary of the NSF workshop on "Scientific Foundations of Qualitative Research" provides a very useful summary of the criteria that define rigorous, high-quality qualitative research in my field. I'm not sure that I can improve much on these, so I will use this memo to focus on three issues that are mentioned in the report but that might be fleshed out more fully. All of these have particular salience for the type of work done by political scientists engaged in qualitative, comparative-historical research, but I think they also apply to other disciplines as well. My points A-C below address the first two questions raised in the workshop charge, while point D deals with the third question.

Indicating the Type of Theoretical Contribution Expected from the Research

The Ragin et al. report rightly pointed out that qualitative research is often conceived and executed with goals in mind that may be quite different from that of other types of work, especially some quantitative statistical work. Not all theoretically informed research is primarily engaged in "theory testing" in a strict sense (certainly not all in the same sense). Some qualitative work is aimed, rather, at other theoretical contributions, such as theory refinement, concept development, or tracing the causal mechanisms that appear to lie behind observed correlations or relationships. As an example, qualitative research in my field, focusing on carefully chosen "critical" cases, has been important in correcting mis-specified theories on the development of certain labor market and welfare state institutions, by showing that prevailing interpretations of very robust statistical correlations essentially had the causal arrows reversed (to put it too crudely, not strong unions → large welfare states and centralized collective bargaining, but rather centralized collective bargaining and large welfare states → strong unions). In order for qualitative research not to be viewed as an "inferior" version of quantitative work (because smaller "n") — and especially so that reviewers can apply the appropriate criteria – it is important for qualitative researchers to articulate explicitly what goals have motivated the conception and design of the research.

Articulating the Logic of Comparison/Case Selection

In comparative historical work in political science (and I presume this applies more broadly too), and partly for reasons hinted at in (1) above, cases are often selected for reasons other than their "representativeness." Since the logic of comparison or case selection is typically crucial to the success of qualitative research projects, it is important that the researcher explicitly address this

56 I draw at some points in this memo on ideas elaborated in more detail in (Thelen 2002) and, to a lesser extent, (Locke and Thelen 1995).

— again, among other reasons, so that the research strategy is transparent for reviewers and so that the appropriate standards are applied. A researcher's choice of cases is frequently guided by the state of current theorizing on a particular subject — sometimes focusing on "critical cases" that allow for direct assessment of received wisdom on causal mechanisms or linkages; sometimes comparing across cases that are typically grouped together in the literature (or seen as very different); sometimes zeroing in specifically on a case or cases that appear anomalous in light of existing theory and that therefore may be fruitful from the perspective of theory development.

In terms of research that is specifically comparative, one of the well known advantages of qualitative research is that it gives the researcher the advantage of close knowledge of and familiarity with the cases under study. Proximity to the empirical cases has the distinct advantage of increasing the probability that the concepts with which the analyst is working capture what he or she is trying to get at, and, especially, that they capture the same thing across all cases under consideration (see especially Collier 1998; Coppedge 1999; also Locke & Thelen 1995; Mahoney & Rueschemeyer 2002: 6-8). In addition to these widely recognized advantages, in some cases, qualitative comparative historical research can lead to wholly new framings of traditional research questions by comparing cases that research designs formulated at a greater distance would not put together or might even see as "non comparable." In such cases, however, establishing equivalency is extremely important and a failure to explicitly articulate the logic of comparison is likely to lead reviewers to misunderstand and therefore discount research proposals.

The general point is that qualitative researchers should take care to articulate explicitly the logic governing their selection of cases, whether or not the research has an explicit comparative dimension.

Attending to Issues of Application/Replication and Specifying the Scope Conditions That Apply to the Claims Being Advanced

Issues of replication and falsifiability are often raised in connection with qualitative work, and so it is important for researchers to address explicitly how the claims advanced in the work can be put "at risk" (or alternatively, how other researchers might "check the work" of their colleagues who employ qualitative methods). In doing so, scholars might keep in mind the several different kinds or levels of application/replication that could be brought to bear, which include but are not limited to:

(1) Same cases, new data. Whereas in some quantitative research, a case can consist of a single "data point," in comparative historical work any given "case" will consist of a multitude of observations. In many ways the strong empirical grounding of claims made in much qualitative research makes putting the findings of that research "at risk" a much more straightforward task than it is for theories that employ highly pliable concepts formulated at great distance from the empirical cases. Other researchers interested in "checking the work" of the comparative-historical scholar can go one of two routes, either "revisit" the same documents and sources as the original researcher, or collect additional observations on the case that could confirm or disconfirm the interpretation or claims being advanced.

(2) Different cases but within the scope conditions stipulated in the original design. Often qualitative research based on close examination of a limited number of cases is designed to apply to a larger number of cases that are seen to be similar in terms of the core causal mechanisms or relationships observed in the smaller sample. In such cases, the theory can be tested against additional observations outside the original cases but within the broader category of phenomena. This requires that the researcher be very explicit on the scope conditions that

determine the boundaries of such applications. What is this a case of? To what class of cases can the theory be meaningfully applied? These scope conditions can be defined either with reference to empirical (for example, geographic or temporal) bounds, or with reference to theory, but either way, being explicit about the "reach" of the theoretical propositions set out will be crucial to ensuring that appropriate standards are applied to the evaluation of the work.

(3) Different cases outside the original scope but employing concepts as originally defined for use inside. It is often the case that qualitative research yields insights that apply far more broadly than for the class of cases for which the research was designed. Recent work on path dependence and policy feedback is an example (e.g., Pierson 1994). Originally designed to explain the resilience of welfare policies in the rich democracies, Pierson identifies causal mechanisms (positive feedback, increasing returns effects) that can be applied more broadly, and the broader application of these concepts yields further insights into the important characteristics of their operation. In such instances, the presence of similar causal mechanisms observed in other cases is not a confirmation of the original (welfare state) theory, but nor is their failure to obtain in some other case a disconfirmation of it. Rather, application of the concepts employed beyond the original case or cases can lead to refinement of those concepts and/or a specification of the general conditions in which such causal mechanisms obtain. In other words, whereas contrary findings in (1) and (2) might be considered disconfirming, contrary findings in (3) would not be strictly speaking disconfirming, though they might well lead to refinement of the concepts and/or theory.

In sum and in general, qualitative researchers should devote explicit attention to questions of scope conditions and appropriate application of concepts employed in their research, so that other researchers can test (and appropriately apply) the claims advanced in the study.

What Areas or Topics Are Most Promising for Investigations Using Qualitative Methods

In my field, the debates on the "absolute" merits of different approaches and methods may be subsiding and giving way to a more constructive mutual engagement process that taps into the relative strengths of different modes of analysis based on the kinds of empirical puzzles scholars are trying to solve. Thus, for example, a number of authors have suggested that formal (mathematical) models are most fruitfully applied in contexts in which the rules and parameters of interaction are established, stable, and well known (e.g., Bates 1997; Geddes 1995). By contrast to this, the strength of a good deal of qualitative comparative historical work is precisely in the leverage it provides on understanding *configurations of institutions* (Katznelson 1997) and over *longer stretches of time* (Pierson 2004) – including where the parameters themselves are changing . In thinking of the specific areas or topics that seem particularly promising for investigations using qualitative methods (question 3 of workshop charge), I would emphasize two: the study of temporal sequences unfolding over time, and the development of theory on institutional origins and evolution. The first of these is in fact a longstanding strength of comparative historical work in sociology and political science, the second defines a newer research frontier that is currently being pursued by scholars associated with a variety of different methodological and theoretical orientations. I'll say a few words about each.

Macro historical processes and political outcomes. Comparative historical scholarship, both the classics and contemporary scholars, has always attached a great deal of importance to issues of sequencing and timing in the analysis of important macro-historical processes (Shefter 1977; Gerschenkron 1962; Lipset & Rokkan 1968; Ertman 1997). A large literature on "critical junctures" of various sorts has probed the significance (for a variety of political outcomes) of the interaction

effects among different processes as they unfold over time, and as they unfold differently (including in a different order) in diverse contexts (e.g., Collier & Collier 1991). The attention to sequencing in historical institutional research is partly motivated by the old truism that in order to establish causality you have to demonstrate not just a correlation between two variables, but also provide some mechanism or theoretical account showing why this linkage exists (Rueschemeyer & Stephens 1997). However, beyond that, the emphasis on timing and sequencing in historical institutional research is also motivated by the insight, borne out in a number of studies (and emphasized especially by Pierson recently) that *when* things happen can itself be extremely important part of the causal story (Pierson 2000). Time series data can be useful for sorting through some of these issues, but qualitative research methods – which specifically focus on processes unfolding over time and the interaction effects among intersecting processes – are explicitly attuned to them as a matter of design. Related to this, macro historical comparative research is well equipped to uncover the deep long-term causal connections between temporally distant "causes" and "effects," thus making a distinctive contribution in political science (at least) where explanations are frequently sought in the immediate temporal vicinity of the outcome to be explained (Pierson 2004: esp 96-102). An example is James Mahoney's explanation for contemporary patterns of development (and underdevelopment) which links these outcomes to colonial legacies (as opposed to other contemporary causal variables such as market reform, etc.).

Institutional genesis, reproduction and change. A good deal of work in political science is organized around the study of how institutions shape political outcomes – with controversy and debate centering less on whether or not institutional factors are significant (most scholars agree they are) than about how best to define institutions and to organize the study of their effects. Many, perhaps most, studies take institutions as "given," and work forward from there to institutional effects (i.e., institutions as independent variables invoked to explain some other outcome). However, a growing number of scholars have begun to turn their attention to issues of institutional creation, reproduction, and change – and on these questions qualitative methods hold special promise. In terms of *institutional origins,* for example, comparative historical research in my field has been employed to identify "critical junctures" – turning points that established important institutional parameters that subsequently shape what is politically possible, even conceivable – thus illuminating aspects of political life that do not emerge through other sorts of analytic strategies or points of departure. In terms of *institutional stability or continuity,* a number of authors already mentioned above have explored the processes of "positive feedback" that account for the stable reproduction of particular rules or arrangements over time (see also Skocpol 1992; and Mahoney 2000). And finally, qualitative research is especially well suited to addressing questions of *institutional evolution and change* and especially modes of change that are incremental but cumulatively transformative (therefore also unfolding in many cases over long periods of time) (see, among others Thelen 2004; Streeck & Thelen 2005). Qualitative research – with its emphasis on process -- has distinct advantages over alternative non-qualitative methods when dealing with all of these general issues.

References

Bates, Robert H. (1997). "Comparative Politics and Rational Choice: A Review Essay." *American Political Science Review,* 91:699-704.

Collier, David. (1998). "Comparative Method in the 1990s." *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 9 (Summer):1-2, 4-5.
Collier, Ruth B., and David Collier. (1991). *Shaping the Political Arena.* Princeton, NJ: Princeton University Press.

Coppedge, Michael. (1999). "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics." *Comparative Politics,* 31:465-476.

Ertman, Thomas. (1997). *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe.* New York: Cambridge University Press.

Geddes, Barbara. (1995). The Uses and Limitations of Rational Choice. In *Latin America in Comparative Perspective: New Approaches to Analysis and Methods,* edited by P. Smith. Boulder, CO: Westview Press.

Gerschenkron, Alexander. (1962). *Economic Backwardness in Historical Perspective.* Cambridge: Harvard University Press.

Katznelson, Ira. (1997). "Structure and Configuration in Comparative Politics." In *Comparative Politics. Rationality, Culture, and Structure,* edited by M. I. Lichbach and A. S. Zuckerman. Cambridge: Cambridge University Press.

Lipset, Seymour Martin, and Stein Rokkan. (1968). "Cleavage Structures, Party Systems, and Voter Alignments". In *Party System and Voter Alignments: Cross-National Perspectives,* edited by S. M. Lipset and S. Rokkan. New York: Free Press.

Locke, Richard M., and Kathleen Thelen. (1995). "Apples and Oranges Revisited: Contextualized comparisons and the Study of Comparative Labor Politics." *Politics & Society, 23*:337-67.

Mahoney, James. (2000). "Path Dependence in Historical Sociology." *Theory and Society,* 29:507-548.

Mahoney, James, and Dietrich Rueschemeyer. (2002). "Comparative-Historical Analysis: Achievements and Agendas." In *Comparative Historical Analysis in the Social Sciences,* edited by J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press.

Pierson, Paul. (1994). *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment.* Cambridge, UK: Cambridge University Press.

Pierson, Paul. (2000). "Not Just What, But When: Timing and Sequences in Political Processes." *Studies in American Political Development,* 14.

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

Rueschemeyer, Dietrich, and John D. Stephens. (1997). "Comparing Historical Sequences: A Powerful Tool for Causal Analysis." *Comparative Social Research,* 17: 55-72.

Shefter, Martin. (1977). "Party and Patronage: Germany, England, and Italy." *Politics & Society, 7*:403-451.

Skocpol, Theda. (1992). *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States.* Cambridge, MA: Belknap.

Streeck, Wolfgang, and Kathleen Thelen, eds. (2005). *Beyond Continuity. Institutional Change in Advanced Political Economies.* Oxford: Oxford University Press.

Thelen, Kathleen. (2002). "The Explanatory Power of Historical Institutionalism." In *Akteure-Mechanismen-Modelle: Zur Theoriefahigkeit makro-sozialer Analysen,* edited by R. Mayntz. Frankfurt/Main: Campus.

Thelen, Kathleen. (2004). *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States and Japan.* New York: Cambridge University Press.

Qualitative Methods in Political Science

Lisa Wedeen
Department of Political Science
The University of Chicago

From the inception of the American Political Science Association in 1903 to the present, there have been repeated attempts within the association to "transform the study of politics into an independent science" (Ross 1991: 288). Despite important variations among positivists and significant disagreements between positivists and non-positivists (including disputes about what "positivism" means), efforts to make political science a science have generally entailed separating facts from values, identifying law-like principles governing political action, formulating hypotheses, and subjecting these hypotheses to empirical tests.

Although standards of rigor are in part a technical matter, subject to debate in any context (from poetry to physics), in political science, arguments about scholarly rigor have generally accompanied efforts to unify the discipline across sub-disciplinary boundaries. A case in point is the exceptionally influential book by Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry* (1994). I think it is fair to say that no book in recent years has been as powerful as this one in authorizing experts and disciplining the discipline. By 2001, over twenty thousand copies had been sold; the book had already been reprinted six times; and five hundred eighteen libraries had purchased it.[57] Insisting that differences in traditions within the discipline were simply *stylistic*, the authors sought to produce a unified epistemological and methodological community, one in which the scientific methods familiar to quantitative researchers would also become the norm in qualitative studies. The unity sought by King, Keohane, and Verba (KKV) was not therefore based on the argument that qualitative work is potentially both non-scientific and legitimate. The claim, rather, was that there is simply no political science worthy of the name that does not conform to the putatively generalizable scientific strictures they defined.

According to KKV, the "best qualitative research" operates with "the same underlying logic of inference" as the one on which quantitative research is based. This logic of inference, whether causal or merely "descriptive," can be made "systematic and scientific," and is "relevant to all research where the goal is to learn facts about the real world" (6). Yet there is little discussion about what constitutes the real world, and the authors concede that not all questions of abiding concern for politics can be covered by the rules of inference. Thus the "real world" seems to become that which is constituted by the "rules of inference." The effect is that some topics are foreclosed for the sake of methodological practices coincident with the authors' specific understandings of science. The belief that such an approach is ontologically, rather than merely provisionally, adequate may signal an unacknowledged metaphysical commitment underlying the book. Certainly such a conviction seems to limit the range of possibilities open for rigorous work in political science. For KKV, "the distinctive characteristic that sets social science apart from casual observation is that social science seeks to arrive at valid inferences by the systematic use of well-established procedures of inquiry (6)." "Good research," for which the authors used "the word 'scientific'" as the "descriptor," is work that adheres to the dictates of explicitly scientific research (7). "Valid inferences" are those established by scientific work. Scientific work assures objectivity.

---

57 Keisha Lindsay supplied this information through a Word CAT Internet database search (June 2001) and through a telephone interview with Eric Rohmann, Sales Director, Princeton University Press, June 11, 2001. The author would also like to thank the participants at the NSF workshop on Qualitative Methods (May 2005), as well as Rohit Goel, Michael Dawson, and Don Reneau.

King, Keohane, and Verba's methodological treatise thereby rested on familiar understandings in the discipline: they assumed not only the intrinsic worth of scientific studies, but they also posited a specific and by no means self-evident understanding of science as a practice based on a clear divide between empirical facts and philosophical values. Like their predecessors, KKV drew a sharp distinction between "what is" and "what ought to be."[58] Questions about issues such as "agency, obligation, legitimacy, citizenship, sovereignty, and the proper relationship between national societies and international politics" were understood to be "philosophical rather than empirical" (KKV 1994: 6). The KKV approach thus not only reproduces the meaningful divide between political science and political theory, locating abstract conceptual concerns outside the domain of proper science, but it also seems to read the manifestly political concerns of theory out of the discipline of political science. Science, following Popper and the behavioralists, required testable, falsifiable hypotheses, an acknowledgement of the tentative nature of findings, and (therefore) an emphasis on methods over results. Worried that the absence of consensus about what science is necessarily entailed disagreement about what constituted good work, the authors attempted not to resolve the underlying philosophical problems raised by Popper et al. but to impose a specific form of scientific rigor on the discipline, at the expense of other rigorous forms of engagement with politics. In this sense, the book fit well with a number of books in the 1980s and 1990s, many of which lamented the divisions within political science and sought to insist on the methodological assumptions of the natural sciences.[59] Although *Designing Social Inquiry* has yet to create the desired consensus, the book was arguably more successful than any other in specifying the terms under which scholarly work would be taken seriously in the field.

Interpretive social science, as a type of qualitative inquiry, received little recognition in *Designing Social Inquiry*, and indeed its methods are rarely taught in qualitative methods seminars in political science more generally. Admittedly, "interpretive social science" is a rubric that can refer to a variety of different epistemological, methodological, and political commitments. The "interpretive turn" is sometimes used as a synonym for the "cultural turn" (Bonnell and Hunt 1999) and at others for "hermeneutics" (Rabinow and Sullivan 1979; 1987; Geertz 1973; 1980.) It sometimes means a commitment among practitioners "to violate the positivist taboo against joining evaluative concerns with descriptions of fact" (Rabinow and Sullivan 1979; 1987), and it sometimes connotes a belief that such a divide is impossible to sustain in practice, so that normative claims and factual statements necessarily infuse one another. Geertz is a practitioner of hermeneutics whose ideas pose an alternative to structuralism; Foucault is often considered to be "beyond structuralism and hermeneutics" (Dreyfus and Rabinow 1983). Both scholars may be reasonably termed interpretivists.

The label may be so elastic as to refer to everything and nothing at the same time. Nevertheless, there are at least four characteristics uniting interpretivists, despite their differences.[60] **First, interpretivist social scientists tend to view knowledge, including scientific knowledge, as historically situated and enmeshed in relationships of power.** Borrowing from Foucault, interpretivists question the "kind of power that is presumed to accompany…science" (Foucault 1972: 84.) They are therefore committed to thinking through the epistemological commitments undergirding the production of law-like principles governing human behavior, which is to say, to dealing with the philosophical questions the scientific approach tends to suppress. **Second, interpretivists are also "constructivists" in the sense that they see the world as socially made**, so that the categories, presuppositions, and classifications that refer to particular phenomena are manufactured rather than

58 For a sophisticated discussion of the distinction between is and ought, see Hanna Fenichel Pitkin's *Wittgenstein and Justice: On the Significance of Ludwig Wittgenstein for Social and Political Thought* (Berkeley: University of California Press, 1993 [1972]).

59 The most obvious text is Gabriel Almond's *A Discipline Divided: Schools and Sects in Political Science* (Newbury Park, CA: Sage Publications, 1990). See also Gary King, *Unifying Political Methodology: The Likelihood Theory of Statistical Inference* (New York: Cambridge University Press, 1989). One of the most oft-cited critiques of *Designing Social Inquiry* is Brady and Collier (eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman and Littlefield, 2004).

60 These four points are excerpted from Lisa Wedeen, "Concepts and Commitments in the Study of Democracy" in *Problems and Methods in the Study of Politics.* Edited by Ian Shapiro, Rogers M. Smith, and Tarek E. Masoud (Cambridge: Cambridge University Press, 2004), p. 284.

natural. There is no such thing as ethnicity or race, for example, outside of the social conditions that make such classifications meaningful. The task of an interpretivist may be, then, to investigate the work these categories do, while accounting for how they come to seem natural when they do. **Third and relatedly, interpretivists tend to eschew the individualist orientation characteristic of rational choice and behaviorist literatures**. Although some interpretivists stress the importance of agentive individuals (e.g., Bourdieu 1977) they do not assume a maximizing, cost-benefit calculator who is unproblematically divorced from actual historical processes. Ideas, beliefs, values, and "preferences" are always embedded in a social world, which is constituted through humans' linguistic, institutional, and practical relations with others (Wedeen 2002). **Fourth, interpretivists are particularly interested in language and other symbolic systems, in what is sometimes termed "culture" in the literature.**

Thus, it might be argued that the interpretivist inclination, rather than seeking to overcome the divergences and disagreements among political scientists, would be to encourage attention to the underlying philosophical issues that make sense of the differences, even while insisting on some shared standards of rigor. Interpretive work, like most work in political science, is subject to evaluation on the basis of such criteria as the logical coherence of the argument being advanced, the cogency of supporting evidence adduced, and the argument's ability to anticipate objections – to take into account alternative explanations and arguments. Interpretivists also insist that social science work should unsettle existing assumptions in ways that are surprising and counterintuitive – inviting us to think in fresh and new ways about issues of perennial concern to political life.

In this light, an interpretivist's appreciation of ambiguity or "complications" need not be confused with what some social scientists might call the "unfalsifiability" of interpretivists' arguments. As I write in my first book, *Ambiguities of Domination: Politics, Rhetoric and Symbols in Contemporary Syria* (1999), my interpretivist account there can be falsified by demonstrating the existence of a regime in which "tired slogans and empty gestures foster allegiance and actually generate people's emotional commitments to the regime" (pp. 152-153). But not all interpretivist work is falsifiable. An analysis of the work qat chews – the leafy stimulant drug Yemenis chew in the afternoons – do in contemporary Yemen, for example, is intended to clarify conceptual puzzlements about "democracy." And case studies of various qualitative stripes might contribute to on-going theoretical work by generating propositions for testing (e.g., that the very fragility of some authoritarian states – with weak capacities to generate national loyalty – may enhance opportunities for widespread political activism and critical, deliberative, public discussion).

Interpretive work in political science is served well by drawing on other disciplines whose practitioners have thought concertedly about interpretation -- ranging from the semiotics of Roland Barthes, through structuralist anthropology, hermeneutics, practice-oriented post-structuralist anthropology, and science studies, to name a few. What a political scientist can bring to bear on those disciplines most familiar with interpretative methods is a fine-grained, rigorous attention to conceptual clarification through the political theory of philosophers such as Ludwig Wittgenstein; an appreciation for the importance of ambiguity in the existing theories of thinkers such as Michel Foucault and Thomas Hobbes (both of whom are being used these days to think about sovereignty); and a particular kind of tacking back and forth between theoretical insights and empirical evidence so that theoretical work illuminates the empirical world and the empirical world also raises questions about our long-standing theoretical presuppositions and findings. Political science is particularly good at asking the "so what?" question. Why should we care about identity-formation in contemporary Yemen (Wedeen, in preparation) or "swanking taxis" in South Africa (Hansen 2006), for example? Coming up with compelling responses can make social science work relevant across disciplinary and sub-disciplinary divides.

Interpretive work focuses primarily on matters related to "culture," but the most promising studies avoid conceiving of culture as a set of sedimented essences inhering in particular groups. In contrast to many political culture approaches, for contemporary interpretivists "culture" does not mean group traits, nor does the term designate a closed symbolic system synechdochic with a particular group. Instead, culture refers to practices of meaning-making – to "semiotic practices." Studying meaning-production entails analyzing the relations between actors' practices (e.g., their work habits, self-policing strategies, knowledge-gathering activities, and leisure patterns) and their systems of signification (language and other symbolic systems). This conceptualization operates in two ways. First, culture as semiotic practices can be thought of as an abstract theoretical category, a lens that focuses on meaning, rather than on, say, prices or votes. It offers a view of political phenomena by paying attention to how and why actors invest them with meaning. Second, this formulation refers to what symbols *do* – how symbols are inscribed in practices that operate to produce observable political effects (Wedeen 2002).[61]

A practice-oriented cultural approach helps us explain issues of abiding concern to political science, such as how rhetoric and symbols generate compliance, how political identities crystallize or change over time, how preferences get generated, and why particular material and status interests are taken for granted, are viewed as valuable, or become available idioms for dissemination and collective action. By paying attention to the ways in which certain meanings become authoritative, while others do not, political scientists can use this practice-oriented concept of culture to help explain why recognizable events or empirical regularities occur. Insofar as studying "culture" refers to investigating practices of meaning-making, interpretive social science gives political science tools for considering questions that are manifestly political, but which have tended to dodge the discipline's purview.

By thinking of meaning construction in terms that emphasize intelligibility, as opposed to deep-seated psychological orientations, a practice-oriented approach avoids confusions that have bedeviled scholarly thinking and generated incommensurable understandings of what culture is across the disciplines. So, for example, Samuel Huntington's notion of culture as a bounded civilization in which essences inhere in particular groups – e.g., Westerners, Muslims – is simply incommensurable with the idea of culture as semiotic practices. These two distinct notions have fundamentally different objects of inquiry. Culture as semiotic practices, I want to argue, has added value because it enables social scientists to examine the historical processes and power relationships that generate a group's "thin coherence" (Sewell 1999) without assuming that coherence a priori. As Rogers Brubaker has pointed out, "violence becomes 'ethnic' (or 'racial' or 'nationalist') through the meanings attributed to it by perpetrators, victims, politicians, […] researchers, relief workers, and others. Such acts of framing and narrative encoding do not simply interpret the violence; they constitute it as ethnic" (Brubaker 2004, 16).[62] An approach that privileges this understanding of meaning-production also allows us to recognize and explain the heterogeneous practices and vigorous communities of argument that exist within putatively coherent nation-states, traditions of piety, etc.

While every activity has a semiotic component, the point here is not to assert that politics must be examined from a semiotic-practical point of view. Whether one does or does not explore processes of meaning-making will be determined by the particular research problem one confronts. A critical understanding of culture as practices of meaning-making can facilitate important insights about politics, enabling political scientists to produce sophisticated causal arguments and to treat forms of evidence that, while manifestly political, most political science approaches tend to overlook.

More generally, qualitative methods (both interpretive and non-interpretive kinds) may be used to think through a variety of research problems. Ethnographic work may be able to give us a subtle

61 For an elaborated version of this argument, see Lisa Wedeen, "Conceptualizing Culture: Possibilities for Political Science," in the *American Political Science Review* (96) 4, December 2002.

62 The ways in which some accounts are "performative," helping to produce what they claim to describe, is treated at length in Lisa Wedeen, *Peripheral Visions: Political Identifications in Unified Yemen*, in preparation. My discussion there is inspired by J.L. Austin, Jacques Derrida, Judith Butler, Pierre Bourdieu, Rogers Brubaker, and Saba Mahmood.

and nuanced comparison of the divergent ways in which Hutus and Tutsis experience their ethnicity in towns and in refugee camps, respectively (Malkki 1995). Historical and comparative analysis, which tends not to be interpretive, may help us understand the origins of elite actors and the constraints they face in "democratic transitions," in industrializing, or in the processes of state-formation (Pierson 2004). Interpretive work might compliment such qualitative studies by questioning what democracy means or by considering how categories associated with democracy get institutionalized over time. Textual analysis can help us analyze how notions of national unity are idealized in official sources, for example, while also enabling us to think about how these discourses operate to produce concrete political effects. Attending carefully to the specific logics of a discourse on piety or on democracy, to name two examples, requires investigating the relationships between the concepts and practices constitutive of a particular "discursive tradition" (Asad 1986; Mahmood 2005). But whereas texts do not actually "talk back," people in unstructured interviews do, thereby enabling us to understand the consequences of particular images or actions, the multiple but nonetheless specifiable ways in which people make sense of their worlds.

References

Almond, Gabriel A. 1990. *A Discipline Divided: Schools and Sect in Political Science.* Newbury Park, CA: Sage Publications.

Asad, Talal. 1986. "The idea of an anthropology of Islam." *Occasional Papers Series*. Washington, DC: Center for Contemporary Arab Studies, Georgetown University.

Bonnell, Victoria E., and Lynn Hunt, eds. 1999. *Beyond the Cultural Turn: New Directions in the Study of Society and Culture*. Berkeley: University of California Press.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Translated by Richard Nice. New York: Cambridge University Press.

Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield.

Brubaker, Rogers. 2004. *Ethnicity Without Groups*. Cambridge, MA: Harvard University Press.

Dreyfus, Hubert L., and Paul Rabinow. 1983. *Michel Foucault: Beyond Structuralism and Hermeneutics*. Chicago: University of Chicago Press.

Foucault, Michel. 1972. *Power/Knowledge: Selected Interviews and Other Writings, 1972*-1977. Translated by Colin Gordon, Leo Marshall, John Mepham, Kate Soper. New York: Pantheon Books.

Geertz, Clifford. 1973. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.

Geertz, Clifford. 1980. *Negara: The Theatre State in Nineteenth-Century Bali*. Princeton: Princeton University Press.

Hansen, Thomas Blom. 2006. "Sounds of Freedom: Music, Taxis, and Racial Imagination in Urban South Africa. *Public Culture*, Vol. 18, no.1, winter, pp. 185-209.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* New York: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Mahmood, Saba. 2005. *Politics of Piety: The Islamic Revival and the Feminist Subject.* Princeton: Princeton University Press.

Malkki, Liisa H. 1995. *Purity and Exile: Violence, Memory, and National Cosmology among Hutu Refugees in Tanzania.* Chicago: University of Chicago Press.

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

Pitkin, Hanna Fenichel. 1993. *Wittgenstein and Justice: On the Significance of Ludwig Wittgenstein for Social and Political Thought*. Berkeley: University of California Press.

Rabinow, Paul, and William M. Sullivan, eds. 1979. *Interpretive Social Science: A Reader*. Berkeley: University of California Press.

Rabinow, Paul, and William M. Sullivan, eds. 1987. *Interpretive Social Science: A Second Look*. Berkeley: University of California Press.

Ross, Dorothy. 1991. *The Origins of American Social Science*. New York: Cambridge University Press.

Sewell, William H., Jr. 1999. "The Concept(s) of Culture." In *Beyond the Cultural Turn: New Directions in the Study of Society and Culture*, eds. Victoria E. Bonnell and Lynn Hunt. Berkeley: University of California Press.

Wedeen, Lisa. 1999. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. Chicago: University of Chicago Press.

Wedeen, Lisa. 2002. "Conceptualizing Culture: Possibilities for Political Science." *American Political Science Review*. 96 (4): 729-744.

Wedeen, Lisa. 2004. "Concepts and Commitments in the Study of Democracy." In *Problems and Methods in the Study of Politics*, eds. Ian Shapiro, Rogers M. Smith, and Tarek E. Masoud. New York: Cambridge University Press.

Wedeen, Lisa. Forthcoming. *PERIPHERAL VISIONS: Politics, Power, and Performance in Yemen.* Chicago: University of Chicago Press.

Appendix 6.

Sociology

Group Report

Defining Qualitative Research Standards in Sociology

In 2003 the NSF Sociology Program convened a workshop on *the Scientific Foundations of Qualitative Research* at which sociologists and other social scientists articulated (1) guidance for both reviewers and investigators on characteristics of a strong qualitative research proposal and (2) recommendations to the Program on how to strengthen qualitative methods in sociology and the social sciences in general. http://www.nsf.gov/pubs/2004/nsf04219/nsf04219.pdf The former reflect the generally accepted standards of rigor for qualitative research used by the NSF Sociology Program in its review process. The latter resulted in a call for research for *Strengthening Qualitative Research through Methodological Innovation and Integration.* http://www.nsf.gov/sbe/ses/soc/sqrmii.jsp

The group revisited the initial discussions and agreed that while all the "standards" might not be appropriate for a particular research project when developing a proposal for submission to the Sociology Program applicants are advised to:

- provide a clear articulation among theory, research questions, design, mode of data collection, and analysis, even if you expect that the actual conditions under which research will be conducted will lead you to modify your plans;

- offer a clear sense of how the project will move from research questions to data collection to data analysis and distinguish clearly between analytic procedures (e.g., categorization) and data analysis;

- present preliminary hypotheses and motivate them by the literature or your knowledge of the case at hand; spell out how the acquisition of contextual knowledge about the topic will lead to greater specification of the hypotheses;

- provide details about the methodological design to be used and anticipate what reviewers likely will want to know;

- detail the standards and methods of data collection (e.g., field notes, interview notes, textual and cultural objects, observation v. participant observation, in-depth interview v. standard interview, etc.) and explain what will constitute data in the study; this is particularly important if you will be using unfamiliar types or sources of data;

- model the planned analysis (e.g., if you plan to do thick description, provide a sample of this in the proposal);

- spell-out the procedures you will use in data analysis (e.g., explain exactly what do you propose to do with your fieldnotes—do not just say that you will use a computerized program (like NUD*IST or Atlas.ti) to do analysis, but explain how and what you will look for in your analysis such as coding categories, archetypal figures, and observed themes;

- explain how you will develop a coding scheme and, insofar as possible, provide a sample of likely coding categories (e.g., provide a sample of text and explain what you would be looking for in your analysis). NSF allows dissertation proposals (but not faculty proposals) to include addenda, so coding schemes and preliminary interview schedules could be included in these proposals;

- spell out tentative hypotheses if aspects of the research design are expected to become clearer as you gain better contextual knowledge of the topic;

- present the research in a manner understandable to all reviewers, including quantitative researchers, that is, the discussion of all methodological issues should be sufficiently transparent to give readers confidence that you know what you are talking about;

- anticipate reviewer objections and respond to them (e.g., if you will not be generalizing beyond your data, explain why; clarify why selecting on a dependent variable is not a problem, if this is the case; explain why bias is not an issue in the study; justify why you are selecting a unique case; explain why snowball sampling is a good strategy for your study);

- provide evidence of the project's feasibility and the your preparedness (e.g., include findings from a pilot study if possible; demonstrate that you can get access to your research site and that you will be able to get the data you need; document your training or experience in conducting this kind of research);

- be clear about the impact of your presence on the research (e.g., if you are riding in a patrol car, how likely is it that you will be able to observe instances of misconduct such as racial profiling by the police; if you are conducting focus groups, how will you deal with the bias that would result from the presence of a dominant, talkative group member?);

- discuss the ethical implications of your work beyond the issues used by IRBs (institutional review boards); how might you ensure that your results are available to those you study?; and

- avoid discussing intricate matters of the philosophy of science or theory that undergird the proposal unless these are directly relevant to the research design.

The group recommended that NSF consider as "standards" in addition to the generally accepted one, the implications of research findings to various constituencies, potential original contributions and other indicators of quality, in evaluating qualitative research. Both investigators and reviewers should contemplate the implications of a study for various constituencies. Specifically, will other researchers or the subject(s) of the research have access to the findings? How clearly has the PI considered the broader ethical considerations of this work? Both should also think about potential original contributions, which are likely to emerge from the qualitative work. Does the project involve unusual research sites, promising juxtapositions of literatures, innovative or intellectually risky ideas or designs? Lastly, proposal should be judged on indicators of quality such as:

- is the project likely to reveal mechanisms and causal forces?
- does the PI provide evidence of the ability to understand the context of the research; does the PI have the necessary cultural fluency or language skills?

- will the project use comparisons or other methods to identify anomalous cases in data analysis?
- are the proposal's procedures transparent?
- does the project aim for an abundance of evidence in its data collection plan?
- is the research design coherent and attentive to alternative explanations, if relevant?
- does the proposal outline a plan to link theory construction and data analysis?

Similarity in Standards and Commonalities with Other Disciplines

Sociology uses the same or similar standards to those used in other disciplines. The group, therefore, agreed that it is important to focus less on the differences among disciplinary paradigms and more on commonalities, especially in terms of standards of evaluation. The need for a clear articulation among theory, research question, design, mode of data collection, and analysis is a common standard. Further, the attention to cultural fluency (including language) and a deep understanding of cultural and social context, commonly used in Anthropology, are useful standards as well for qualitative research in Sociology.

Promising Areas of Qualitative Research in Sociology

In Sociology most topics are amenable to both qualitative and quantitative research, but there are some especially promising areas for qualitative strategies and data.:

Projects for Qualitative Research

- highly complex social structures, processes, and interactions
- studies of the mechanisms underlying causal processes, especially over time
- naturally occurring processes and phenomena of social life
- studies that focus on questions of 'how' and the connection between the 'how's' and the 'why's'
- the use of in-depth interviews to clarify findings from survey research

Topics for Qualitative Research

- studies of scientific research and evaluation
- manifestation of globalization at the micro level, including as units of analysis, cities, cultural practices, families, interpersonal relations, urban labor markets, and gender relations
- studies of how race, class, gender, age intersect and play out in various everyday contexts and with respect to everyday concerns, such as health
- the mechanisms that underlie patterns of inequality and social inclusion/exclusion
- consequences of war and social conflict on communities and for the construction and dissolution of collective and personal identities
- conceptions of equality and inequality
- religious beliefs and political participation in America and elsewhere

Resources to Strength Qualitative Research

The group offered a number "programmatic" suggestions that NSF might consider, including those that involve building on existing resources as well as ones where new resources are needed to further strengthen qualitative research.

- More funding for qualitative research and more publicity of NSF's commitment to fund such work. The NSF workshops at the annual American Sociological Association meetings are excellent and should be continued.

- Expand funding of pilot, exploratory, and small grants through its on-going support for the Fund for the Advancement of the Discipline (FAD) of the American Sociological Association (ASA). The availability of FAD funding should be advertised on the NSF website to more widely disseminate information about this NSF funding opportunity.

- Work with with ASA to develop pre-dissertation funding opportunities for students conducting qualitative research.

- Consider funding release time for PIs who need large blocks of time, which is an inherent part and cost of much research that utilize qualitative approaches.

- Support activities that would develop and strengthen qualitative research training such as ones that enhance the teaching of qualitative data analysis, such as summer schools in qualitative sociology similar to those held a The Interuniversity Consortium for Political and Social Research (ICPSR) or, with ASA, could sponsor qualitative training workshops.

- Fund more research on knowledge production and evaluation practices to study how people think about what leads to the production of excellent work, thus help us better understand the "science" of research quality and excellence.

- Create a qualitative data bank from preexisting NSF-sponsored research to use for teaching qualitative methods.

- Explore ways to help qualitative researchers work successfully with their home institutional review boards to establish reasonable and workable standards for human subjects review for qualitative work.

- Support an institute for advanced studies in qualitative research or a discipline-wide teaching workshop in qualitative methods.

- Collaborate with Harvard University's Murray Center, the University of Michigan's ICPSR and other data depositories to develop procedures to archive qualitative data while assuring confidentiality.

Information, Communication, and Outreach

Communicating qualitative research standards to both NSF reviewers and investigators are an important step. NSF should communicate a set of evaluation standards for qualitative proposals to panelists and include panelists who understand the challenge of evaluating qualitative research. A short list of do's and do not's in evaluating qualitative work would be helpful. NSF could also recognize the value of qualitative training by graduate programs by asking applicants and panelists to demonstrate, where appropriate, formal qualitative research training. In order to make qualitative research more widely available NSF could explore ways to heighten confidentiality protection for qualitative data. The development of procedures to make qualitative data available to other researchers and students while protecting the identity and confidentiality of subjects has the potential to transform qualitative data analysis and dissemination.

Sociology Papers Presented by Workshop Participants

Qualitative Research Standards of Rigor and Sociology & How They Might be Communicated

Kathleen M. Blee
Department of Sociology
University of Pittsburgh

These suggestions are intended for research proposals that use ethnographic, socio-historical, or in-depth interviewing methods, those with which I am most familiar. Some of these are based on ideas developed by scholars engaged in qualitative research in the field of nursing. I do not cite individual works in the text, but a list of relevant articles in nursing is appended.

Standards of Rigor

In ethnographic, socio-historical, and in-depth qualitative interviewing studies in sociology, the standards for rigor involve a number of aspects of the research design, including:

- The project should have a clear *focus*. Ideally, it should be possible to summarize the goals of (or the questions driving) the study in 1-2 sentences and the significance of these goals/questions in another 1-2 sentences.
- The research should have a clear *starting point*, but generally, *not* be clear about how it will proceed throughout to the endpoint. The starting point for research should include a description of: (i) sources of data and access to these; (ii) the process whereby particular data units (persons, events, interactions, etc) will be identified, selected/sampled, and acquired; and, (iii) if the selection of data is influenced by pragmatic considerations (e.g, membership in a group), an acknowledgment of that and a discussion of its advantages and disadvantages. In most qualitative designs using ethnographic, socio-historical, or in-depth interviewing, it should be clear that initial data collection/analysis may require alterations in data collection, sampling, identification of variables or subjects, and/or research design, so these should not be specified in the proposal in a rigid fashion beyond the initial stages of the project.

- There should be a clear sense of what will constitute *data* in this study or a sense of how that will be determined in the process of the study.
- There should be a clear sense of how the researcher will look for *patterns* in the data, while preserving a sense of the complexity of social life. Is s/he looking for numerical frequency, foundational status, commonness, or other features, and how does that fit the overall goal of the research design? There should be a statement about how negative cases will be sought and what status they will have in the design as well as what other measures will be used to increase the researcher's accountability to the data.
- There should be explicit (possibly even standard, within the project) means by which data will be *assessed* to help ensure that: (i) all data are considered,; (ii) more spectacular acts/events/ individuals are not overly stressed; and (iii) data that are different from the pattern are not discounted without a clear rationale for doing so.
- The design should be *sufficient* to produce the claims that are intended from the study.
- For most (but not all) projects, there should be a clear distinction between *analytic procedures*

and *data analysis.* Generally, lists or frequencies of themes or categories (in interviews), reflexivity (in ethnography), or event occurrences (in socio-historical studies) should be presented as analytic procedures rather than analysis unless these involve new interpretations.

- For many (but not all) projects, analysis (including 'thick description') should consider *relationships* among concepts or a sense of their origin, development, construction, etc. rather than presentations of individual concepts, dimensions, themes, or categories.
- There should be more than a formalistic or bureaucratic assessment of the *ethical issues* involved in the study, particularly concerning fundamental ethical and political concerns that arise in research with human subjects, beyond and different from those covered in standard protocols of Institutional Review Boards (IRBs).

Communication of Standards of Rigor

- NSF could create a template for the assessment of rigor in qualitative research proposals across the social sciences and include that on its website.
- NSF could create a template that sensitizes reviewers (especially in more highly quantitative social science fields) on how to assess a research plan that looks different from a deductive study, in which, for example, there is not likely to be a dependent variable or a hypothesis.
- NSF could have available on its website several mock 'bad' and mock 'good' qualitative proposals, with detailed reviewer commentaries on the research design.
- NSF could create additional incentives (in the form of supporting teaching buy-outs) for research projects that involve teams of qualitative researchers from multiple disciplines studying a common phenomenon or site.

Areas of divergence between Sociology other Social Sciences

In Sociology – and perhaps even more so in Political Science and Economics – description is viewed as an inadequate outcome of research, even in qualitative studies. This may be less the case in Anthropology, History, and some interdisciplinary social sciences. Surrounding the undervaluing of description in these social sciences (as causes, consequences, or parallel developments):

- Description is rarely taught as a methodological goal in these disciplines.
- There is little or no consensus or explicit standards for what constitutes rigorous or high-quality description (as opposed to just a mass of detail) and how this could be ascertained or achieved.
- There has been some shift – at least in qualitative sociology – away from methods (such as ethnographies) that are likely to yield rich description of a social context and toward methods (like interviewing) that are less likely to do so.
- Abstract renderings of data tend to be valued by reviewers and the discipline, relative to interpretations and analyses that remain closer to the data.
- Analytic techniques and methods (such as narrative analysis) are imported into research proposals to avoid the negative evaluation that a project is 'simply descriptive' when the study's aim might be better served by excellent description.

Perhaps through high-profile workshops at NSF, there could be an effort to create a multidisciplinary set of standards or protocols of excellence for description as an outcome of some social scientific projects. Drawing on the experience of some of the 'hard sciences' in which rigorous

description is both valued and well established (e.g., biology, astronomy, etc) might be useful for social sciences.

Promising Areas of Research Using Qualitative Methods

Qualitative methods have proven useful for examining social phenomena as emergent processes, for uncovering causal connections, and for delineating sequencing of actions, events, or understandings over time. Encouraging more qualitative projects that collect data over time would be expensive due to the intense time commitment necessary by investigators in many kinds of qualitative data collection and analysis projects, but would have good payoff in terms of developing a better understanding of the causation and sequencing of various factors in social life.

Research Design and Methods Issues of Qualitative Research

The volume of data generated by many kinds of qualitative research designs can prove overwhelming, especially to new and student researchers. Data management software like NVivo7 and others are helpful in sorting these data but require a tremendous amount of time at the front end for coding.

The cost of interview or focus group transcription is very high, which encourages researchers (especially students) to use only very small samples. Better in-depth interviewing protocols that combine standardized and computer-ready information gathering procedures with less structured methods would be helpful here, as will be likely advances in voice recognition software that can be used in interviewing situations.

There are not good protocols for the reporting of findings in qualitative research. Although qualitative research will never be as straightforward as is the case in quantitative research due to epistemological and methodological issues, including, importantly, the critical role of context in qualitative studies, it is often difficult, as Sandelowski and Barroso (2002) note, to "find the findings" in qualitative studies or to discern how these fit with findings of other studies on your topic. This impedes the accumulation of knowledge from qualitative research designs and the use of qualitative research findings in quantitative research.

Tools, Training, Data, Research Design, and Infrastructure Needs

- Training students in qualitative analysis (rather than just qualitative data collection) is needed to move beyond all-too common statement in research proposals that themes and categories of analysis will somehow 'emerge' during the research process.
- A consistent policy among university and other Institutional Review Boards about the status of various forms of qualitative data collection and, optimally, a blanket exemption for some forms of very low risk and non-invasive data collection similar to that recently granted for oral history projects. A strong message about these issues from the social science units of NSF might help, especially with NSF-funded projects. The language on the NSF website about ethnographic observation is a great step in this direction, but could be expanded to include other types of low-risk qualitative work.
- Extended time away from teaching for qualitative scholars to develop deep descriptions of the contexts of their research projects.

- Financial support for collaborative or team approaches in ethnographic or other time-consuming observational methods, including funds that would allow students to work on these projects in collaboration with (rather than in place of) faculty observers.

- A set of large, internet accessible, and high-quality qualitative data bases collected through different qualitative data collection techniques – including interview and focus group transcripts, ethnographic field notes, photographs and video of social interactions in a variety of contexts – to be used for teaching qualitative analysis to students. NSF might fund a special initiative to get data from a variety of previously-funded qualitative projects compiled in a format that would be useable for this purpose.

References

Belgrave, Linda L, Diane Zablotsky, and Mary Ann Guadagno (2002). "How Do We Talk to Each Other? Writing Qualitative Research for Quantitative Readers." *Qualitative Health Research,* 12: 1427-39.

Morse, Janice. (2003). "A Review Committee's Guide for Evaluating Qualitative Proposals." *Qualitative Health Research* 13: 833-51.

Penrod, Janice. (2003). "Getting funded: Writing a Successful Qualitative Small-Project Proposal." *Qualitative Health Research,* 13: 821-32.

Sandelowski, Margarete. (2000). "Whatever Happened to Qualitative Description?" *Research in Nursing & Health*. 23: 334-40.

Sandelowski, Margarete. (2001). "Real Qualitative Researchers Do Not Count: The Use of Numbers in Qualitative Research. [Review] [29 refs]." *Research in Nursing & Health*. 24: 230-40.

Sandelowski, Margarete, and Julie Barroso. (2002). "Finding the Findings in Qualitative Studies." *Journal of Nursing Scholarship*. 34: 213-9.

Sandelowski, Margarete and Julie Barroso. (2003). "Writing the Proposal for a Qualitative Research Methodology Project." *Qualitative Health Research,* 13: 781-820.

Multidisciplinary Standards for Systematic Qualitative Research

Wendy Griswold
Department of Sociology
Northwestern University

My discipline is sociology, and for the most part there are no standards for qualitative work, at least, none in the scientific sense. I see three reasons for this. First, much of qualitative sociology – and here I am referring mostly to ethnography and to cultural analysis – has taken its cue from the humanities, whereby producing and illustrating a compelling theory is sufficient. (Sometimes this move is couched in a poststructuralist critique of positivism, although this no longer seems to be mandatory.) Second, an additional source of the lack of rigor is sociology's disciplinary roots in the progressive social movements of the early 20th century and in journalism, neither of which employ the scientific method to advance their claims (Robert Ezra Park is an exemplary and influential figure in this regard). Third, a new element that mitigates against rigor is sociology's small-but-increasing attention to areas of the world outside of the West or Japan. In these emerging areas of investigation, data of all sorts – archival, textual, quantitative – is often hard to find, and issues of comparison and interpretation become especially tricky. This may encourage a less systematic and more impressionistic research.

So how do we recognize good work? A few rule-of-thumb criteria are commonly applied:

- *Comparison.* Looking at two to six cases of similar phenomenon and focusing on how they differ, or drawing on interviews and observations of lots of people and organizing them along a couple of dimensions (e.g. four or six cells) is preferred. (Examples: Binder 2002; Lamont 2002)
- *Immersion.* Thick description leading to the reader's sense of saturation, of being within the phenomenological world and understanding its meaning systems, is preferred. Accounts of worlds we do not know much about are better than accounts of familiar worlds. (Examples: Jackall 2005; Duneier 1999; Fine 1996)
- *Transcendence.* Powerful theory, illustrated by vivid quotations or observations, is preferred. This theory draws on data, but in some ways transcends it; the theory itself organizes the data, and there is no sense of testing hypotheses. (Examples: Swidler 2003; Alexander 2005; Wacquant 2003)
- *Piling up.* Amassing huge quantities of data, often including some qualitative data backing up an essentially qualitative analysis, is preferred. Such work is often to some extent historical, but in contrast to historians, sociologists aim for breadth more than depth. (Examples: Fischer 1992; Zelizer 1997; Griswold 2000)

Areas or topics most promising for investigations using qualitative methods

First the obligatory caveat: Most areas and topics benefit from a combination of qualitative and quantitative methods. Relegating any area of knowledge to one style of research or another is a mistake; methodological multiplicity is a virtue.

Now to address the question as asked. Qualitative methods are advantageous for:

- Investigations where the analytic categories are not known. This is especially true for minority or non-mainstream groups. For example, a survey investigating the spread of HIV might attempt to categorize people as heterosexual or homosexual. However, if the survey designers did not know that some urban African American men employ have a concept of being "on the down low" (heterosexual men having occasional homosexual relations), then a quantitative analysis would miss that category of sexual behavior. Qualitative research has a better capacity to elicit previously unknown social categories.
- Investigations of what people value, what makes them happy, what bothers them, and why. Individuals and groups very in their preferences and in the weight they give these preferences. Survey data is notoriously clumsy at addressing this variation, often making too much or too little of the variables under consideration rather than understanding them as part of a cultural complex.
- Investigations of everyday practices, including for example the maintenance of cultural boundaries and the spatial and symbolic dimensions of social inequalities.
- Investigations of organizational, community, or small-group cultures.
- Investigations of cultural objects: ideological, religious, aesthetic, material, or literary.

Most pressing issues of research design and methods facing qualitative researcher projects

Designing the research so that one's initial assumptions may turn out to be incorrect is one of the most pressing issues of research design.. Some form of what I have called provisional, provincial positivism – "if I'm right about this, we should see X, and if we do not find X, I'm probably not right" – is essential to guard against sociologists' tendency to confirm what they already believe.

Area of promising qualitative research most likely to foster multidisciplinary projects

What has been called the "new urban studies," which draws heavily on anthropology, landscape-ecological history, and cultural studies as well as economics, political science, and geography, is a natural area for qualitative analysis. The work of neo-urbanists like Richard Sennett, Eric Klinenberg, and Sharon Zukin exemplifies this.

Tools, training, data, research design, and infrastructure for conducting qualitative research

I think "training" suggests the wrong model. I prefer reading. People need to read more and read more deeply in the historical and comparative literature on their area of investigation, and on its artistic and literary and journalistic representations. Sociologists (and members of other disciplines) are too inclined to confine their reading to a narrow approach to/definition of the area in question. The training model comes from the natural sciences in which there is a well-defined problem and a well-specified literature on it. In contrast, many important social questions are initially ill-defined, so the researcher must cast a wide and non-specific net.

References

Alexander, Jeffrey C. (2005). *The Meanings of Social Life: A Cultural Anthology.* Oxford: Oxford University Press.

Binder, Amy J. (2002). *Contentious Curricula: Afrocentrism and Creationism in American Public Schools.* Princeton: Princeton University Press.

Duneier, Mitchell. (1999). *Sidewalk.* New York: Farrar, Straus, and Giroux.

Fine, Gary A. (1996). *Kitchens: The Culture of Restaurant Work.* Berkeley: University of California Press.

Fischer, Claude (1992). *America Calling: A Social History of the Telephone to 1940.* Berkeley: University of California Press.

Griswold, Wendy. (2000). *Bearing Witness: Readers, Writers, and the Novel in Nigeria.* Princeton: Princeton University Press.

Jackall, Robert. (2005). *Wild Cowboys: Urban Marauders & the Forces of Order.* Cambridge: Harvard University Press.

Klinenberg, Eric. (2003). *Heat Wave: A Social Autopsy of Disaster in Chicago.* Chicago: University of Chicago Press.

Lamont, Michèle. (2002). *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration.* Cambridge: Harvard University Press.

Sennett, Richard. (1991). *The Conscience of the Eye: The Design and Social Life of Cities.* New York: Knopf

Swidler, Ann. (2003). *Talk of Love: How Culture Matters.* Chicago: University of Chicago Press.

Wacquant, Loic. (2003). *Body & Soul: Notebooks of an Apprentice Boxer.* Oxford: Oxford University Press.

Zelizer, Viviana A. (1997). *The Social Meaning of Money: Pin Money, Paychecks, Poor Relief, and Other Currencies.* Princeton: Princeton University Press.

Zukin, Sharon. (1993). *Landscapes of Power: From Detroit to Disney World.* Berkeley: University of California Press.

Preliminary Remarks on Qualitative Research Standards in Different Disciplines

Michèle Lamont
Department of Sociology
Harvard University

Debates about the evaluation of qualitative research within each of the fields represented at the workshop are contingent on the ecological environment in which they develop. In political science, much of the recent writing on the question has been framed in response to the challenges raised by King, Keohane, and Verba's (1994) very influential *Designing Social Inquiry*. In anthropology, Clifford and Marcus's (1986) *Writing Culture* has generated a cycle of collective reflection on how the identity and the position of the researcher affects her work, which has framed the discussion about reliability and validity in terms almost incompatible with how the question is framed in psychometrics for instance: instead of bracketing the identity of the researcher, the challenge is to fully understand its impact on research. In sociology, a multimethod discipline par excellence, we are now going through a phase where more scholars are concerned with the similarities between the evaluation of qualitative and quantitative research, as a growing number of students are being trained for "multi-methods." Yet, the repercussions of a long-lasting disciplinary split between quantoids and qualtoids continue to be felt.

Observing the impact of broader disciplinary ecologies on disciplinary conversations about the standards of qualitative work helps us understand some of the possibilities and limitations of our task. Because qualitative method is practiced in very different environments in anthropology, criminology/ legal studies, political science, and sociology, we should focus less on our differences than on the *greatest common denominators shared across disciplines.* We have to concentrate on the areas where standards of evaluation intersect. Furthermore, our charge is to reflect on standards of evaluation not for all qualitative research conducted in the social sciences, but only for the standards to be applied to proposals submitted to NSF. It goes without saying that many researchers may not want to conform to these standards.

My own insights into the topic at hand come from my experience as a qualitative sociologist, from teaching graduate seminars in qualitative research, and from conducting research on definitions of quality used in funding panels in the social sciences and the humanities.[63] This research focuses on how panelists draw the line between the winners and the losers, and on disciplinary differences in how scholars define quality across disciplines. My forthcoming book *Cream Rising: Finding and Defining Excellence in the Social Sciences and the Humanities* is addressing the formal and informal criteria of evaluation used in multidisciplinary panels. As such, it takes as its object *knowledge evaluation practices*. I consider the standards of evaluation of qualitative research to be a subset of this broader issue. Thus, at times, this memo is concerned with what I think is right, or should be right, and at other times, with what I believe others believe is right.

63 I conducted interviews with panelists serving five different fellowship competitions and twelve funding panels in the social sciences and the humanities over a period of two years. The funding competitions were held by the Social Science Research Council (SSRC), the American Council of Learned Societies (ACLS), the Woodrow Wilson National Fellowship Foundation (WWNFF), a Society of Fellows at a top research university, and a foundation in the social sciences. These competitions were chosen because they cover a wide range of disciplines, and because they are all highly prestigious. While the SSRC and the WWNFF competitions are open to the social sciences and the humanities, the ACLS supports research in the humanities and in humanities-related social sciences. The Society of Fellows supports work across a range of fields, whereas the anonymous foundation supports work in the social sciences only. Moreover, the SSRC and the WWNFF programs provide support for graduate students, whereas the ACLS holds distinct competitions for assistant, associate, and full professors. The Society of Fellows provides fellowships to recent Ph.D.'s only, and the anonymous social science foundation supports research at all ranks. We did not identify any bias in the evaluation of the work of graduate students.
A total of 81 interviews with panel members in charge of final deliberations were conducted. This total includes 66 interviews with 49 different panel members (17 panelists were interviewed twice, because they served on panels for the two years that the study lasted). Fifteen additional interviews were conducted with relevant program officers and panel chairpersons for each panel, who provided details about what had happened during the panel deliberation in the absence of direct observation.

Standards of Rigor in Qualitative Research

The interviews I have conducted with panelists suggest that these standards vary depending on whether panelists adopt a comprehensive epistemological standpoint influenced by a Weberian *Verstehen* approach to knowledge production, a constructivist approach to knowledge production, or a positivist approach to knowledge production. These three standpoints are typically used by qualitative researchers in the social sciences, with the comprehensive approach being much more popular than the other two. While the first group emphasizes the importance of insight, meaning-making, and attentiveness to context, the second group shares this concern, but also emphasizes how the identity of the researcher shapes and motivates research. For its part, the third group emphasizes generalizability and, in some cases, falsification (Mallard, Lamont, and Guetzkow forthcoming).

Despite these differences, in all cases, the fit between theory, methods, and data and the justification of the research procedures, including data analysis, are stressed as crucial standards that proposals should meet. Providing a cogent account of how various decisions concerning the research design are made, and of how these tie with the theoretical motivation of the project, are a *sine qua non*. Thus, researchers of all stripes value making available to the evaluators the information needed to assess the work accomplished. Being explicit about how one goes about producing research, and thus allowing others to reproduce results if desired, is of the essence (for a similar point, see Susan Silbey's memo in this volume).

My study of funding panels, which in most cases considered proposals for qualitative research projects, also strongly suggests that most members of panels agree that they share many standards of evaluation. This is confirmed by the fact that in most competitions, a consensus around a sizable proportion of the strongest and weakest proposals emerged in the rankings submitted by panelists prior to deliberation. This indicates that if panelists are not always able to articulate precisely what defines quality, many concur that "we recognize it when we see it." But, what is this "it"? Attentiveness to details, clarity, and tightness of the connection between theory, data, and methods are crucial, as is of course originality, defined largely as daringness, ability to frame a problem differently, or ability to open new vistas and ask new questions (on originality, see Guetzkow, Lamont, and Mallard 2004).

I noted above that in sociology, a growing number of departments abide by the motto of multimethod training, thereby sharing a common assumption that there is not one good method, only good questions, and that a good researcher should demonstrate an ability to mobilize the tools best suited to address a given question. This translates into what I perceive to be an important subcultural shift within the field: quantitative and qualitative researchers are moving away from thinking of themselves as distinct and even incompatible breeds. The best researchers are the ones who are able to do several things reasonably well. This shift also means that the standards of rigor in qualitative and quantitative research are increasingly convergent, as researchers use related mental templates as they move across topics. However, qualitative research does not necessarily come increasingly to resemble quantitative research (following a template proposed by King, Keohane, and Verba 1994). Instead, there is growing awareness of the theoretical/empirical back-and-forth needed in qualitative and quantitative research alike, a perspective advocated for instance by Ragin (1994; on the logical problems addressed by different methods, see also Stinchcombe 2005). Instead of positing that quantitative research shows us the route for high-quality research, qualitative researchers are reflecting on the distinctive requirements to be met by qualitative research. In the study of racial inequality, for instance, scholars are becoming increasingly aware that the identity of the interviewer should not be bracketed, and that there are distinct advantages associated with having a group's insiders and outsiders interview someone, since the respondent's presentation of self is unavoidably shaped by whom she is talking to (on this point, see Lamont 2004a). The goal is not necessarily to produce generalizable knowledge, but to produce research

that sheds new light on social phenomena and that adds specificity and complexity to our understanding. As David Snow (2004) proposed, the goal often is to provide theoretical refinement and extension to earlier formulations, as opposed to formulating falsifiable hypotheses. These developments are entirely compatible with new trends in other fields–the growing literature on process tracing in political science, for instance (George and Bennett 2005; see also the exchange in *Sociological Methodology* regarding Gorski 2004).

The predominant discourse on qualitative methods in the social sciences continues to stress the absence of shared standards–as suggested by some of the memos prepared by participants in this workshop, and indeed, by the charge given by NSF to our multidisciplinary group. Nevertheless, we have to remember that qualitative researchers routinely make collective decisions concerning academic products–dissertations, articles, books, fellowship applications, etc. While there are always divergences concerning the details of the evaluation, many organizational and procedural mechanisms encourage the development of a common matrix. Evaluators learn to bracket incompatible standards, and they come to share cognitive schemas and feeling rules concerning how they should behave to achieve results that are viewed as fair and legitimate (they should be methodologically pluralists, show mutual respect, and be inclusive, for instance, as shown in Mallard, Lamont, and Guetzkow forthcoming; Lamont forthcoming). Panelists develop a common understanding about where the center of gravity of a group lies within the first few hours of a meeting, just as we did in the first hours of our workshop as we labored to understand where others stood. For their part, program officers encourage panelists to formulate their arguments in a language that makes sense to others and orient group dynamics so as to avoid voting and instead produce consensual decisions. The social processes that make collective decisions and shared standards possible remain largely misunderstood, and one of the challenges ahead of us remains to study empirically the extent of cognitive consensus around standards of evaluation for qualitative research, and the social and cultural mechanisms that produce them.

Communicating Standards for Rigor

A) Within sociology

NSF could do much more to socialize applicants and panelists about what common standards for the evaluation of qualitative research are–all panelists are presumed competent to evaluate by the mere fact of having been invited to serve, yet we know very little about how academics of all stripes learn to evaluate scholarship, given that this task is never taught explicitly.

The memos prepared for this workshop propose several promising steps toward improving the socialization of applicants and evaluators alike–posting templates of good and bad projects on NSF's website and offering national workshops to improve training in qualitative research, for instance. One more action that is likely to help the diffusion of more consistent standards is to increase the number of qualitative research proposals funded by NSF. As potential applicants discover that NSF is indeed funding qualitative research, they will be more likely to send proposals themselves, and to encourage their students to do the same. By writing proposals and receiving feedback, graduate students will be socialized into thinking more systematically about what defines a strong qualitative proposal that meets NSF standards. At the same time, NSF will develop greater expertise in the evaluation of such projects. I gather that NSF's interest in funding more qualitative research, and in standards for qualitative work, grew from an unanticipated and continuous sizable increase in the number of qualitative dissertation improvement grant proposals in recent years.

B) Across Disciplines

Again, my research suggests that there is already quite a bit of overlap in standards across the social sciences, in that what we call the "comprehensive standpoint," which corresponds to Weber (1949)'s *Verstehen*, seems to be privileged by the majority of social scientists whom I have interviewed for my project on funding panels. Of course, the consensus found among evaluators serving on competitions at the Social Science Research Council and the Woodrow Wilson Fellowship Foundation may be very different than those found at NSF, where panels tend to be uni-disciplinary and perhaps wedded to the more strongly institutionalized understandings of how each discipline differs from others. One could probably decide on a range of standards that are shared by social scientists across all the fields represented here, at least for the purpose of evaluation of NSF proposals. We can agree that proposals should be very explicit about the type of data the research will draw on and how the researcher will go about collecting the data. The areas of divergence—about, for instance, whether reflexivity is essential to good research—may best be left undiscussed. Again, a focus on the greatest common denominators can be much more easily reached than the adoption of a common set of principles that should be applied unequivocally to all fields (concerning generalizability or how to establish validity, for instance).

Pressing Issues of Qualitative Research and Methods

We need to gather systematic information about how qualitative and quantitative research are similar and different from one another by interviewing researchers concerning how they understand their work, e.g., whether they think knowledge "cumulates," whether they produce "findings" (and what kinds of findings), and what they think gives value to various types of research, including specific recent examples of celebrated pieces of scholarship. The literature abounds with normative statements concerning the technical procedures to follow to produce high-quality research, but we know very little about the "laboratory life" of the social sciences, along the lines of Latour (1993), for instance.[64] This is an essential next step. As Abend (2006) puts it, the time is ripe for an empirical sociology of epistemology. In particular we need to address very systematically how top-notch qualitative researchers understand their distinctive contributions to the social sciences, and what gave value to the most widely cited books and articles published in a field. To take only one example, we need to explore what difference was made by Arlie Hochschild's (1979) introduction of the concepts of "emotion work" and "feeling rules" in gender relations, or Ann Swidler's (1986) "cultural toolkit" metaphor. Did these concepts allow for more knowledge accumulation and generalizability? If not, what *is* their contribution? An analytical device for capturing phenomena that had gone unnoticed previously? How can we explain the appropriation of these tools by such a large number of researchers in recent years? What difference did they make to the knowledge we produce? I suspect that much of the qualitative research that has been most influential is research that has *generated new insights and opened new vistas*, not necessarily research that excels if assessed by the standards of validity, replicability, and generalizability. This would also certainly be the case for Erving Goffman's (1963) *Stigma*, Clifford Geertz's (1973) analysis of the Balinese cock fight, or Albert Hirshman's (1964) *Voice, Exit, and Loyalty*.

Promising Areas of Qualitative Research and Foster Multidisciplinary Projects

Most topics can be studied using qualitative or quantitative methods, and different aspects of a phenomenon are brought to light by the use of different techniques. Even our understanding of meaning making, a social process that is most frequently approached using qualitative, interpretive techniques, can be enriched by the use of more accurate measurement techniques (see, for instance, Jepperson and Swidler 1994; Mohr 1998).

64 Latour (1993) shows that when presenting their research to their peers in publication, biologists use a "reductionist" rhetoric that provides a linear description of the research process according to which the complexity of empirical reality is broken down into elementary units amenable to analytical treatment and deductive hypothesis testing. This occurs even though their process of discovery often is non-linear and involves induction and feedback loops as well as deduction.

All social science topics are amenable to qualitative analysis. However, we can identify areas of growth and vitality within each discipline, areas toward which large numbers of young people are moving. Some of these areas or topics are particularly conducive to qualitative research. This is the case for cultural and economic sociology, or the study of social movements, for instance (see Lamont 2004b). One could imagine targeting such vital areas as starting points for building networks that would connect young people across a number of fields. The model for multidisciplinary work that is now being put in place by the Social Science and Humanities Research Council of Canada, building on the research program model of the Canadian Institute for Advanced Studies, is a case in point (see in particular the "successful societies program" at www.ciar.ca). These programs bring together researchers who share a research interest but have very different yet complementary expertise. The large "networks of excellence" put in place by the EU function on similar premises and have been extremely successful, changing the face of European social science research (notably in the field of immigration; see Patricio 2004). They have been extremely successful at generating enormous dynamism across large multidisciplinary networks of scholars working on related topics.

Acknowledging the multifaceted character of the social world, which will lead us to acknowledge the need for a diversity of research techniques and approaches? This in turn will lead to a greater appreciation of the specific contributions that can be made by the various approaches, and to recognizing that different standards should be used to evaluate different types of knowledge. Such standards should be conceived as complementary rather than exclusive of one another.

References

Abend, Gabi. 2006. "Styles of Sociological Thought: Sociologies, Epistemologies, and the Mexican and U.S. Quests for Truth." *Sociological Theory* 24 (1): 1-41.

Clifford, James, and George Marcus, ed. (1986). *Writing Culture: The Politics and Poetics of Ethnography.* Berkeley, CA: University of California Press.

Geertz, Clifford. (1973). *The Interpretation of Culture*. New York: Basic Books.

George, Alexander L. and Andrew Bennett. (2005). "Process Tracing and Historical Explanation." Pp. 205-232 in *Case Studies and Theory Development in the Social Sciences.* Cambridge, MA: MIT Press.

Goffman, Erving. (1963). *Stigma*. Englewood Cliffs, NJ: Prentice Hall.

Gorski, Philip S. (2004). "The Poverty of Deductivism: A Constructive Realist Model of Sociological Explanation." *Sociological Methodology*, 34: 1-33.

Guetzkow, Joshua, Michèle Lamont, and Gregoire Mallard. (2004). "What is Originality in the Social Sciences and the Humanities?" *American Sociological Review* 69: 190-212.

Hirschman, Albert. (1964). *Exit, Voice and Loyalty*. Cambridge, MA: Harvard University Press.

Hochschild, Arlie. (1979). "Emotion Work, Feeling Rules and Social Structure." *American Journal of Sociology*. 85: 551-75.

Jepperson, Ron and Ann Swidler. (1994). "What Properties of Culture Should We Measure?" *Poetics* 22: 359-71.

King, Gary, Robert O. Keohane, and Sydney Verba. (1994). *Designing Social Inquiry. Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Lamont, Michèle. (2004a). "A Life of Hard but Justified Choices: Interviewing Across (too) Many Divides," *Researching Race and Racism*. Martin Bulmer and John Solomos (eds.). London: Routledge.

Lamont, Michèle. (2004b). "The Theory Section and Theory Satellites." *Perspectives. Newsletter of the ASA Theory Section*. 27 (1) (January). 5 pp.

Lamont, Michèle. (Forthcoming). *Cream Rising: Finding and Defining Excellence in the Social Sciences and the Humanities*. Cambridge, MA: Harvard University Press.

Latour, Bruno. (1993). "Portrait d'un biologiste en capitaliste sauvage." Pp.100-130. In *Petites leçons de sociologie des sciences.* Paris: Editions Points Seuil Sciences.

Mallard, Grégoire, Michèle Lamont, and Joshua Guetzkow, (forthcoming). "Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review." *Science, Technology, and Human Values.*

Mohr, John W. (1998). "Measuring Meaning Structures." *Annual Review of Sociology* 24: 345-70.

Patricio, Therese. (2004). *Towards an International Strategy with European Research Organizations*. Ottawa: Social Science and Humanities Research Council of Canada.

Ragin, Charles C. (1994). *Constructing Social Research*. Thousand Oaks, CA: Pine Forge Press.

Silbey, Susan. (2004). "Designing Qualitative Research Projects." Pp. 121-126 in *Workshop on Scientific Foundations of Qualitative Research*, edited by Charles C. Ragin, Joane Nagel, and Patricia White (eds.) Arlington, VA: National Science Foundation.

Snow, David A. (2004). "Thoughts on Alternative Pathways to Theoretical Developments: Theory Generation, Extension, and Refinement." Pp. 133-136 in *Workshop on Scientific Foundations of Qualitative Research*, edited by Charles C. Ragin, Joane Nagel, and Patricia White (eds.) Arlington, VA: National Science Foundation.

Stinchcombe, Arthur L. (2005). *The Logic of Social Research*. Chicago: University of Chicago Press.

Swidler, Ann. (1986). "Culture in Action: Symbols and Strategies." *American Sociological Review* 51: 273-86.

Weber, Max.([1905, 1991] 1949). *Max Weber on the Methodology of the Social Sciences.* Edward A. Shils and Henry A. Finch (trans. and ed.) Glencoe, Ill: Free Press.

Conceptions of a Former Program Officer

Joane Nagel
Department of Sociology
University of Kansas
NSF Sociology Program, 2002-2004

From 2000-2004 I served for two years as an NSF Sociology panelist and for two years as NSF Sociology Program Director. During this time I read colleagues' proposals in Sociology, Political Science, Law & Social Sciences, Anthropology, and other social science disciplines, and I read colleagues' reviews of proposals in these areas. Among the several hundred proposals that crossed my screen in those four years, I observed some disciplinary differences in approach (more qualitative methods in anthropology than political science, more modeling in sociology and political science than in anthropology, more cross-disciplinary uses of literatures and methodologies in Law & Social Sciences proposals). Quite often, however, I was struck by similarities among reviewers in identifying strengths and weaknesses in proposals and, often, by the reactions of PIs (Principal Investigators) to those reviews. Below are some of my observations as they bear on the conduct of qualitative research and the goal of identifying standards across disciplines. My comments are organized around three themes: the methodological imperative, the question of theory, and the logic of research design.

The Methodological Imperative

It is a common social science assertion that the research question should dictate the methodology and data used to answer it. My experience, however, is that researchers tend to be wedded to a particular research technique (surveys, ethnography, archival document analysis, interviews, conversation analysis), and that they look around for a research topic on which to ply their chosen methodology: where can I conduct field research, what documents are available for me to examine, what can I count, whom can I listen to, interview, observe? As a result, in sociology, for example, we tend to train "survey researchers," "ethnographers," "conversation analysts," "network analysts," or "comparative-historical" sociologists who toil in different subfields most of their careers and who embrace a relatively fixed set of beliefs about the superiority/inferiority of various methodologies and data sources. A methodology, thus, can become a way of life.

The tendency for social scientists to specialize in particular methodologies leads to the development of disciplinary subcultures that sometimes take on the characteristics of sects—complete with doctrines, founders, believers, converts, heretics, secret codes, initiation rites, revealed wisdoms, sacred rituals: the graduate statistics course sequence, the fieldwork experience, the event history analysis workshop, the archive, grounded theory, significance tests. Methodological specializations produce important and useful expertise, but they sometimes (not always) can create advocates, defenders of the faith, witch hunters, and crusaders for and against particular ways of conducting research. Extreme methodological zealotry (for one's own) and bigotry (against those of others) is relatively rare, more often methodological specializations are marriages of convenience, comfortable arrangements that researchers relax into and cannot imagine working without. Researchers seldom change sects.

In Sociology there is tolerance, even respect, for a variety of research methods, but there still tends to be career specialization in particular methodologies. This can make it difficult to establish standards for, say, qualitative research proposals across a variety of techniques, such as ethnography or

content analysis or in-depth interviews. How many cases (observations, settings, articles, interviews) are enough? How do we know if the observations are accurate or typical? What is the validity or comprehensiveness of coding categories? Is generalizability even a goal? Did the interviewer bias the responses? We cannot necessarily presume unanimity among researchers working within and especially across methodological traditions, not to mention disciplinary differences in methodologies (e.g., ethnography practiced by anthropologists compared to sociologists).

Since the development and dissemination of standards for designing and evaluating qualitative proposals is an important goal for social science, it might be most expedient to recognize rather than resist the institutionalization of methodological differences. The mobilization of methodological practitioners by category may be the most organizationally expedient means for articulating qualitative standards and training future social scientists. Such a strategy would be to organize gatherings of ethnographers to suggest strategies for strengthening ethnographic research, meetings of archival researchers to recommend best practices for archival research, or workshops for interviewers to develop techniques for constructing and conducting oral histories and interviews. The general utility of such segregated endeavors would depend in part on shared epistemologies and research goals, in particular on agreement about the question of theory and the logic of research design.

The Question of Theory

Although different disciplines sometimes seem to have different answers to the question of theory—what is it, where is its proper place in the research process, what constitutes a satisfactory explanation, plausible accounting, convincing argument—virtually all NSF program panels whose deliberations I observed and whose reviews I read expressed interest in answers to questions of why, when, or under what conditions. The question of theory often was answered differently by reviewers and researchers. Reviewers frequently looked for theory in proposals; researchers seldom gave them what they wanted. I found both quantitative and qualitative research proposals typically failed to offer even the most modest theoretical contribution. Most researchers seemed satisfied to identify an "interesting" phenomenon or process and content to offer a means of documenting it—often phrased as "looking at"—whether by counting it, observing it, or talking to it.

The question of theory—to do or not to do—was especially often debated by qualitative researchers. Although I did not find this to be the case in practice, quantitative projects, data, and researchers generally are presumed to be "testing" theories and "predicting" outcomes whereas qualitative projects, data, and researchers are presumed to be "generating" theories and "observing" outcomes. Despite these presumed relationships to theory and hypothesis testing, few projects we reviewed (and many that we funded) while I was at NSF were embarked on systematic hypothesis generation or evaluation. A main complaint of reviewers across disciplines about both qualitative and quantitative projects was the proposal's low level of theoretical development and the project's limited potential for conceptual contribution. Interestingly, a main complaint of PIs across disciplines about both quantitative and qualitative projects was that reviewers were asking for theory. PIs often argued that funding for data collection or analysis alone was perfectly justifiable since their topic was so important, and anyway, they argued, theory either was present in the proposal (i.e., simply was invisible to our obviously blind reviewers) or theory would emerge out of the data (i.e., easily would materialize through the magic of empiricism).

There is certainly room in social science for various styles of research: descriptive, policy, evaluation, model building, theory evaluating. I would argue, however, that the National Science Foundation cannot and should not fund all of them. NSF should specialize in the latter—theory development and evaluation with evidence. Research that is not designed to adjudicate among

explanations, produce new theoretical understandings, or contribute to the conceptual development of a argument, process, phenomenon, should not be funded by NSF. This means that reviewers, program officers, and colleagues need to press researchers to push their work a step (or two or three) further toward richer theoretical framing and production. Standards for qualitative (and quantitative) research should emphatically respond to the question of theory with a clear answer: YES, projects must make a theoretical as well as an empirical contribution.

The Logic of Research Design

In the thousands of email messages I sent as NSF program officer, I often repeated myself. So, I started saving messages that I found myself composing over and over again. Below is one that I frequently sent in response to a research précis sent by a PI planning to write a full proposal or a researcher asking if a project idea was appropriate for the Sociology Program:

> Your project sounds very interesting and might be suitable for submission to the Sociology Program. Successful proposals are those that pose an important and interesting research question, situate the planned project within ongoing and major research in the area, enter into a dialogue with the relevant literature in Sociology and related fields, make a conceptual contribution or advance theory, contain a clear and detailed plan for obtaining and analyzing data, and outline a research design that is falsifiable—one that allows you to be wrong as well as right.

It was that last recommendation—what Karl Popper referred to as "falsifiability," that generated the most questions from my email correspondents. What exactly did I mean by a falsifiable research design, and what did one look like?

Except for experimental researchers, the closest most of us come in our research to designing falsifiable projects is to undertake a systematic search for negative evidence—what are the available alternative explanations for what we are trying explain, and what evidence can we find that these alternative explanations are implausible. When I was studying the dramatic increase in the American Indian population in the U.S. Census (a 72% increase from 1970-1980), I looked for evidence of the usual demographic or technical reasons for such an increase: declining death rates, increasing birthrates, immigration, changes in the measurement of "Indian." When none of these accounted for the growth, I offered my own explanation (increasing Indian ethnic self-identification) and offered an explanation for why it had occurred (federal policies, cultural politics, ethnic mobilization).

When we push ourselves and our students to devise ways to ask and answer research questions in ways that make them even modestly falsifiable, we have a much better chance of improving the theoretical quality of social science research. Just as it is important to resist qualitative researchers' protestations about polluting research with theoretical expectations in advance of data collection, it is important to stress the feasibility of making a theoretical contribution a goal of all research. Challenging researchers to struggle to imagine a falsifiable research design helps them to embark on one of the most direct routes to theoretically and empirically rigorous research.

Two Promising Ways to Advance Qualitative Research

A few years ago a particularly vexing exchange occurred in the *American Journal of Sociology* among ethnographers Loic Wacquant, Eli Anderson, Mitch Duneier, and Katherine Newman. Wacquant did not like their politics, perspectives, or presumptions, and they did not like his politics, pomposity, or presumptuousness. Wacquant's complaints centered mainly on what he saw as a conservative apologia imbedded in the three very well-reviewed and well-received books written by Anderson, Duneier, and Newman, and they defended their work by responding to his criticisms. I sympathize with Wacquant's

targets having been savaged myself in the past by a supercilious critic who called me naïve and my work unsophisticated. But what vexed me most about this "dialogue" was that it centered almost exclusively on ideology and very little on data.

As I noted above, there is room in social science for all kinds of research, and I guess, for all kinds of reviews, debates, and criticisms. NSF can't force researchers to focus on the facts. But we can make it possible for them to do so. Neither Anderson nor Duneier nor Newman nor Wacquant had access to any of the data in the others' books beyond what was published. (They might not have asked anyway—only once have I ever been asked for my data which I shared since their gathering was funded by NSF) It is important is that all data be available—not only quantitative data, but qualitative data as well. This is, in my opinion, the most pressing issue that faces the development of rigorous qualitative research. It is an area in which NSF could have a major impact in two ways: technical assistance and funding. All NSF-funded projects require that data be made available to other researchers. This is not currently possible for many PIs since the procedures and resources needed to make it possible to share data are not available. There are many unresolved problems associated with sharing qualitative data: how to guarantee confidentiality, make data sharing feasible, fund the preparation of data for publication.

Science depends on replication, evaluation, reinterpretation of empirical findings as well as arguments and agreement about the internal and external validity of data collection regimens, analytical strategies, and interpretative frameworks. Progress and the evolution of knowledge—what Thomas Kuhn referred to as "scientific revolutions"—depend on both new ideas and the accumulation of empirical findings that challenge prevailing theories. Social theory can be enlivened as researchers engage in battles of words over bias or politics. I would like to see, however, more arguments about facts. Social knowledge and the credibility of social science cannot advance by debates staged in the absence of evidence. What is needed for informed debates and the advancement of knowledge is evidence that everyone has access to and can evaluate and analyze. Providing technical assistance and funding for qualitative data sharing are important ways that NSF can advance qualitative research.

My second recommendation for strengthening qualitative research is to encourage the design and incorporation of qualitative research methods courses in all graduate programs. Since this involves science education, perhaps support for the development of curricula could be funded in partnership with NSF's Education and Human Resources Directorate. This workshop and the previous workshop and report on qualitative methods sponsored by NSF's Social, Behavioral, and Economic Sciences Directorate are steps in that direction. What is needed next is a means of institutionalizing further the pursuit of rigor in qualitative research. Graduate education is one way to achieve this, another might be through the establishment of a center for advanced study in qualitative research methods

Lost In Translation:
How Not to Make Qualitative Research More Scientific

Mario Luis Small
Department of Sociology
University of Chicago

Introduction

One version of an old joke among Spanish-speaking immigrants tells of Pablo running into his friend María, who is with a group of friends. Everyone speaks Spanish. Pablo, who has been in the U.S. for 10 years, says, "María, introdúceme a tus amigos." The joke is that the phrase does not mean "introduce me to your friends"; it means something closer to "insert your friends inside me." The right phrase is "preséntame a tus amigos," but Pablo, accustomed to the English language, has gotten his languages mixed up. The joke is funny because everyone knows what Pablo meant to say. In other settings, however, the problems of translation can lead to much worse than embarrassing misunderstandings.

Methods of scientific inquiry are languages to the extent that they constitute systems of thought, with terms that have specific meanings and ways of framing problems that make sense only within the system. Most quantitative researchers employ the language of frequentist or classical statistics; qualitative researchers often employ the language of participant observation or the different language of in-depth interviewing. As Howard Becker insists, all methods fundamentally seek the same things (that arguments are backed by data, that procedures can in theory be repeated, etc.); however, the languages with which they say those things are significantly different, and some statements only make sense in some languages.[65]

If methods are languages, then the most important issue facing qualitative researchers---especially those concerned about the science of their work---is translation. This is especially a problem for researchers who study topics, such as neighborhood poverty, where both quantitative and qualitative research is necessary. Many social scientists in recent years have rightly attempted to bridge the gaps between qualitative and quantitative thinking. But many of these attempts have involved making qualitative research, for which there are fewer agreed-upon rules, come closer to matching the logic of inquiry of classical statistics. Thus, qualitative researchers encourage their students to make sure their samples are "representative" or "unbiased," and quantitative ones warn against "selecting on the dependent variable." I believe that in doing this we are forcing words into systems of thought where they do not belong, and that this will exacerbate, rather than improve, our problems of communication.

I hope to make this problem clear by focusing on two examples of how ethnographers concerned about science attempt to make their work "generalizable." My own work is in the fields of inequality and urban poverty, and the translation problems I discuss refer to those between qualitative and quantitative researchers in inequality, simply because this is the work I know best. Some of my work is qualitative and some of it is quantitative; I write as an interpreter hoping to increase the quality of translation, not as a chastiser of one method or another, anymore than one would chastise Spanish for not being English.

65 In this light, it is worth noting that even quantitative researchers have different languages. Classical or frequentist statistics has a common set of terms (central limit theorem, null hypothesis, t-statistic) tied to a way of framing problems different from that of Bayesian statistics, whose own set of terms (prior distribution, posterior distribution, non-informative prior) is tied to its way of framing problems.

First Example: In-Depth Interview

Jane is writing her second-year paper on the attitudes of working-class African-Americans about immigration. She wants to conduct in-depth interviews to capture the nuances of these attitudes, and is planning to interview 35 respondents. Jane worries, "Will my findings be generalizable to the wider population?"

Her adviser, a qualitative researcher, recommends finding a city with a large working class African-American population, obtaining a telephone directory, and randomly sampling people from it. He knows to expect, at best a 50% response rate, so he recommends contacting 100 people selected at random. Jane follows the plan, and miraculously all phone numbers are valid and everyone answers the phone. Of the 100 respondents, 60 hang up on her, 40 agree to an interview, and 35 follow through with it. (Interviewers recognize that, for some populations, these are wildly optimistic figures.) She conducts 35 high-quality 2-hour interviews that delve deeply into attitudes about immigration, uncovering subtle causal relationships among attitudes, experience with discrimination, gender, and Southern origins, and she happily completes her paper. Since her method mirrors that of Lamont's (1992) well regarded *Money, Manners, and Morals*, she is confident in the "generalizability" of her findings about the black working class.

The problem is that under no statistical definition of generalizability can the responses of those 35 individuals be considered to reflect reliably the conditions of the African-American working class. In fact, a quantitative researcher's confidence in Jane's estimates would be just about the same had she simply gone to any neighborhood in the city, and interviewed the heads of households of the first 35 houses on Main Street. It is tempting to think that the first sample is significantly better because it is "more random" but it hardly a statistical improvement.

There are two reasons for this. First, the sample has an inbuilt and unaccounted for *bias*.[66] Jane only interviewed the 35% of respondents who were polite enough to talk to her, friendly enough to make an appointment based on a cold-call from a stranger, and extroverted enough to share their feelings with this stranger for 2 hours. It is very likely that these people have systematically different attitudes about others, including immigrants, than non-respondents. Since we do not know anything about those who did not respond (they hung up), we have no way of adjusting the inferences we obtained from the 35 respondents. In addition, since we do not know anything about working class blacks in other cities, we do not know if, had she had 100% response rates, respondents in her city were typical or atypical of the black working class.

Second, regardless of how it was selected, the sample is *too small* to make confident predictions about complex relationships in the population of working-class blacks at large. Many students ask, "How many people do I need to interview for my findings to be generalizable?" The answer depends on the distribution of the variables of interest, whether the students want to describe distributions (e.g., proportion Democrat) or present causal relationships (e.g., whether Republicans will have stronger anti-immigrant attitudes than Democrats), and how many variables are involved, among other things. King, Keohane, and Verba (1994:213) provide a formula, based on standard statistical assumptions. The short answer, however, is that rarely will students have enough well-selected in-depth interview respondents that their findings about subtle causal relationships involving multiple variables will be statistically generalizable to a large national population. For that, one needs a survey.

Suppose Jane only wanted to know how many working class blacks are pro-immigration reform (one Y/N question); and she wanted to be 95% confident that the average in her sample matched the average in the population at large within +/- 5 percentage points; and that the population of working-

66 For a discussion of these issues from researchers aimed at bridging the qualitative/quantitative divide, see King, Keohane, and Verba (1994:63ff).

class blacks in the U.S. were 2,000,000 people (for large populations, the exact size does not matter very much). In this case, she would need 385 respondents.[67] If Jane narrowed her focus, and only wanted to be confident about the 1,000 working-class blacks in one city, she would need 278.

Some qualitative researchers prefer to ignore these issues and refer to studies such as Jane's as "representative," but the truth is that in doing so qualitative research is simply adopting words without adopting their meaning.

The natural question is whether, having acknowledged this, it is still not better for Jane to have picked her respondents "at random" (in quotation marks because her final sample is not statistically random due to high non-response) than engaging in some other non-random selection strategy. Not always. Consider, for example, sampling for range (Weiss 1994). Suppose Jane suspected strongly that gay and lesbian respondents would be more sympathetic to immigrants. Even a truly random sample would have yielded, at best, 3 or 4 gay or lesbian respondents out of 35, of which 1 or 2, at best, would reveal this to her. This would leave her no room to examine this question. In these circumstances, Jane would be better served designing her study to include a large, pre-determined number of gay and lesbian respondents, even if this meant finding them through non-random means, such as organizations. For many questions of interest to interview-based sociologists, sampling for range is a more effective strategy.

Even in circumstances where researchers are not seeking a particular and small population, random is not necessarily better. Snow-ball sampling, for example, involves asking respondents to recommend other respondents. This has the well-known problem that respondents will tend to be in-network members. Because of this "bias," some researchers are reluctant to recommend this method over random sampling. But snow-balling almost always leads to higher response rates, since people are less reluctant to speak to strangers when they are sent from a trusted source. So, which is worse---the "bias" from in-network sampling or the "bias" from low response rates in random selection? Neither is; which method to employ depends only on the objectives of the project.

Second Example: Neighborhood Study

There is a similar problem in participant observation research aimed at dealing with large-n questions. Bill wants to study how neighborhood poverty affects out of wedlock births, by conducting an in-depth ethnography of a single high-poverty neighborhood. His main concern is the set of mechanisms underlying this process, but he wants to make sure his findings are generalizable to all poor neighborhoods. Thus, he does what King, Keohane, and Verba (1994:67-68) might do:

> For example, we could first select our community very carefully in order to make sure that it is especially representative of the rest of the country…. We might ask a few residents or look at newspaper reports to see whether it was an average community or whether some nonsystematic factor had caused the observation to be atypical…. This would be the most difficult part of the case-study estimator, and we would need to be very careful that bias does not creep in. Once we are reasonably confident that bias is minimized, we could focus on increasing efficiency. To do this, we might spend many weeks in the community conducting numerous separate studies…[68]

---

67 The formula is $n = (Z^2 * p * (1-p))/C^2$, where Z is the Z value (1.96 for a 95% confidence level); C is the confidence interval, expressed as a decimal (in this case .05); and p is the percentage of people who are expected to be, in this case, pro-reform. We assume .5, the most conservative assumption. If 51% are pro- and 49% are anti-reform, the room for error is high, so a large sample is needed; if 90% were pro-reform one could get by on a much smaller sample of 139. There are dozens of sample size calculators on the web, where one can manipulate the assumptions. For example, www.raosoft.com/samplesize.html.

68 In this passage, the authors were discussing much broader issues, so this selection does not do justice to their book. The purpose here is not to produce a full-fledged critique of the authors' (in many ways excellent) book. Rather, it is to show the pitfalls of this particular way of thinking about case study selection, which the authors do share with many others.

Bill looks to the census, finds a neighborhood that is 40% poor, 60% black, with 80% of the households female headed and (he discovers at the site) most streets littered and covered in graffiti, all of which seem to accord with his definition of a "representative" poor neighborhood. Bill conducts his study, and finds that the high level of poverty has made residents distrustful of each other. This distrust, he finds, makes the women unwilling to marry the fathers of their children. Since his neighborhood was representative, Bill is confident that neighborhood poverty increases out of wedlock births in poor neighborhoods at large through the mechanism of lower trust.

The problem with this way of thinking is that, no matter how Bill selected his single neighborhood it will never be truly representative of poor neighborhoods. The neighborhood's conditions may happen to match the traits that, from the census, one knows to be at the mean. But, as Frankfort-Nachmias and Nachmias (2000:167) write, "a sample is considered representative if the analyses made using the sampling units produce results similar to those that would be obtained had the entire population been analyzed." No "sample" of a single neighborhood can match this criterion.

Even obtaining copious and very detailed information on the neighborhood (a generally sensible recommendation by King, Keohane, and Verba [1994]) does not change this fact. Suppose that instead of neighborhoods we were speaking of individuals, and we selected one person with the characteristics of the average American: a married 37-year old female with a high school education who earned $35,038 last year.[69] We interviewed this female for 2 hours about her opinions on the admission of Turkey into the European Union. How confident would we be that her thoughts accurately reflected those of the average American? A scientist would have no confidence, and interviewing her for 20 or 200 additional hours would not change this.

Bill does not have a "sample" of 1; he has a single case study. Suppose that Bill had chosen a neighborhood with a 40% poverty rate but with no garbage or graffiti and a unique architectural design due to the influence of a mayor interested in promoting architecture in the city. It is tempting to think that inferences based on the second case would be less statistically generalizable but, based on a sample of 1, they are neither more nor less so. As before, one could ask if there is any harm in going for the statistics-inspired "random" or "average" strategy. Sometimes there is. Suppose the mayor in the second case also had a radical and unique policy whereby mothers received significantly higher rent subsidies plus $1,000 per child for a college fund if they married before the birth of their second child. This rare case would suddenly present Bill an exceptional opportunity to examine the relationship among high poverty, policy, and out of wedlock births in ways that cases that happen to be at the mean might not.[70] In case studies, rare cases are often precisely what the researcher wants (Yin 2002). In all case studies, though, selection must be made primarily on substance.

Give Up?

My purpose here is not to argue that ethnographic or interview-based methods are destined to be unscientific. On the contrary, I strongly believe that in-depth interviewing and participant observation constitute two of the purest *empirical* methods there are, which is why I rely on them in my work. (Statistical surveys rely on abstractions of the world into pre-determined variables, and thus are inherently once-removed from empirical reality.) My objective is to encourage qualitative researchers to produce scientific work based on their own language, not that of others.

---

69 The median age for males and females is 37; more individuals are married than never married, widowed, or divorced; among persons 25 or older, more are high school graduates or graduates with some college than not high school graduates, college graduates, or persons with advanced degrees; $35,038 is the median earnings for individuals for the year 2002. See Section 1, Population, of the *Statistical Abstract of the United States*, http://www.census.gov/prod/2004pubs/04statab/pop.pdf .

70 These arguments are discussed further in the concluding chapter of my (2004) *Villa Victoria: The Transformation of Social Capital in a Boston Barrio.*

Consider Lamont's (1992) aforementioned book, a study of 160 upper-middle class men in France and the United States. I think Lamont's book is one of the most methodologically sophisticated interview-based studies in recent years. However, I do not believe, as others have commented, that it is sophisticated because "she had a representative sample." The study's response rate was very low, between 42% and 58%, by liberal estimates[71] (Lamont 1992: 218). In addition, the samples are small, only 80 individuals in each country (40 in each site). This is common among field-base studies, and not a problem given the arguments the book makes. But pretending it is truly representative only detracts from the true strengths of the work, and encourages young scholars (like Jane) to focus on making qualitative research more quantitative instead of on improving the way in which they handle qualitative research. The methodological sophistication of the book comes from the sensitivity of the interview process; Lamont's ability to interpret the meaning of respondents' statements within their cultural contexts; her use of a comparative model to sharpen her concepts; her judicious use of both semi-structured interviews, which allow findings to emerge inductively, and a structured survey, which provides comparative data across the cases; her thoughtful selection of research sites (Lamont 1992: Appendix II); and her effective use of these data to tell a compelling story.[72]

One Alternative: Cases, Not Samples

Behind the desperate search for "representative" qualitative data in Bill and Jane's projects is the assumption that if one cannot make statistical statements about the distribution of a variable, one is not engaging in science. I believe this is false. Consider psychological experiments. Most of these are conducted on small and highly unrepresentative samples of college students at large research universities. Yet a recent Nobel was awarded for precisely this type of work.

One way to think about alternative conceptions of scientifically rigorous qualitative work is adopting Yin's (2002) distinction between *case study* logic and *sampling* logic. Yin's work is on case studies, but I believe it is applicable to in-depth interview-based studies, which can be seen, rather than as small-*sample* studies, as multiple-*case* studies. In what follows, I am extrapolating from his work. I cannot do justice to his work in these few pages, but one example should suffice.

Sampling and case study logic approaches are different and fully independent ways of approaching data. In a sampling model, the number of cases is predetermined; the sample is meant to be representative; all individuals should have equal (or known) probability of selection; and all units should be subject to exactly the same questionnaire. In a case model, the number of cases is unknown until the study is completed; the collection of cases is, by design, not representative; each individual has its own probability of selection; and different people have different questionnaires. Case study logic is critical when asking *how* and *why* questions, with which a sampling logic has greater difficulty.

An example from a different method (experiments) will show the fruitfulness of a case study approach. Alfonse conducts an experiment in which one group of black and white students at Berkeley is told they will receive an IQ test and another is told nothing. Both complete the same test, and blacks in the first group do much worse than whites, while those in the second do as well as whites in their group. Alfonse concludes the fear of fulfilling a stereotype about low IQs among blacks is at play.[73] He then does two things, literal and theoretical replication (Yin 2002). With a colleague at Duke, he repeats the experiment among Duke undergraduates (literal replication); back at Berkeley, he repeats it, but using men and women instead of blacks and whites (theoretical replication). If the theory is right,

71 As Lamont writes, the figures "do not include potential respondents who did not provide the information necessary to determine whether they qualified or not" for the study (Lamont 1992:285). Thus, the figures could overstate the response rate.

72 To be clear, I do not think she was *mistaken* in employing a random sampling strategy. The point is that, if we were to judge it by the (inappropriate) standards of statistical generalizability, the sample is no better than many other alternatives, neither of which would fare very well. One cannot expect high response rates when conducting in-depth interviews regarding personal issues for 2 hours.

73 This example is (very) loosely based on the work of Claude Steele and colleagues, at Stanford University.

it should work for anyone, not just blacks and whites. Some results confirm his findings; others do not. Then he tries it among Asians and whites, and among issues other than IQ, and on more campuses, and with high school students and seniors, and on an on. Slowly, as the number of experiments increases, his confidence that his theory is right begins to increase. Eventually, every new experiment contributes very little new knowledge, such that the 89th experiment, with immigrant Asians and Russians in a low-income high school, shows exactly what he expected. At this point, he has attained saturation, and he stops.

Alfonse has just conducted (after many years) a type of multiple-case study. Notice that at no point did Alfonse conduct a random sample of any kind. On the contrary, the characteristics of respondents in every experiment were deliberately chosen.

I suggest that this approach may be used to think about in-depth interview studies. The key is to think about every individual as a single experiment. Jane, without knowing how many respondents she will eventually interview, interviews one. The person recounted experiencing discrimination from Latino immigrants when she was a child, thus developing anti-immigrant sentiments and favoring immigration reform. From the interview, Jane theorizes that blacks who have been discriminated against by Latino immigrants will favor immigration reform. She then searches for blacks who report discrimination from Latinos (literal replication), as well as those who have not experienced it (theoretical replication) and those who experienced discrimination from Russian immigrants (theoretical replication). Importantly, she alters each new interview to make sure to include increasingly refined questions about different aspects of discrimination. She does this over and over. Her last interviews are longer than the first, and they include many more subtle variations on the way one experiences discrimination. Eventually, each new interview is telling her very little she had not already heard about the relationship between discrimination and immigrant attitudes. She has attained saturation.[74]

Jane's method violated nearly all of the tenets of (frequentist) sampling logic. Her group of respondents is not representative; each respondent received a slightly different questionnaire; there was no attempt to minimize statistical bias. Thus, Jane can make no statement about the distribution of attitudes. She cannot report accurately that 25% of working class blacks favor immigration reform, just as Alfonse would not report that 80% of blacks are susceptible to stereotype threat, or that, since 75% of the experiments confirmed his theory, his theory is right 75% of the time (this would be wrong on many, many counts). However, we would have the same confidence in her findings as we do in Alfonse's statements that stereotype threat reduces performance. Jane's work would be as scientific as Alfonse's, even though neither of them can make distributional statements.[75]

Conclusion

Adopting a case-based logic is *not* the only way to conduct ethnographic research that is scientifically rigorous. The point is that it is possible to conduct rigorous research without employing the assumptions of classical statistics in any way. (In fact, the method described above, in its reliance on revising assumptions based on new information conducted during the study, bears some parallels to Bayesian statistics.) To claim that studies such as Bill's are "generalizable" is to adopt terms while ignoring their meanings. It is to mistake "insert your friends inside me" for effective communication. The strengths of qualitative work come from understanding *how* and *why*, not understanding *how many*, and improving this work should mean improving the reliability of its answers to how and why questions. For qualitative researchers to attempt to make their work statistically representative is to engage in a

74 These descriptions of the research process are stylized, as all of them are forced to be. In real life, Jane would have interviewed 10 people before any semblance of a story might have emerged. She then would have limited the scope of her study and her questions, to prevent continuing to interview indefinitely.

75 Of course, if Jane had simply selected her 35 cases as she had before, she would not be able to make statements about distributions either.

losing race, one in which those who have large samples, by design, will always win. It is the equivalent of evaluating success in one language on the basis of the grammar and vocabulary of another. In science, many tongues are better than one.

References

Frankfort-Nachmias, Chava and David Nachmias. (2000). *Research Methods in the Social Sciences*. 6th edition. New York: Worth Publishers.

King, Gary, Robert O. Keohane, Sidney Verba. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. New Jersey: Princeton University Press.

Lamont, Michèle. (1992). *Money, Morals, & Manners: The Culture of the French and American Upper-Middle Class*. Chicago: University of Chicago Press.

Small, Mario Luis. (2004). *Villa Victoria: The Transformation of Social Capital in a Boston Barrio*. Chicago: University of Chicago Press.

Weiss, Robert. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York: Free Press.

Yin, Robert. (2002). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage.

Coming Out from Under the Ethnographic Interview

Alford A. Young, Jr.
Department of Sociology and
Center for Afro American and African Studies
University of Michigan

Introduction

My principal method for sociological research is the ethnographic interview. My research has largely entered on urban-based, low-income African American men and on African American social scientists and humanists who conduct research on, and teach about, the African American experience. In recent decades, ethnographic interviewing has been embraced rather strongly by researchers in various sub-fields in the discipline of sociology (including, but not restricted to culture, ethnic and race relations, organizations, family, and education). Sociologists interested in the study of culture and in race and ethnic relations seem to have led this charge. My work places me in both camps. Put most simply, I interview people about their life experiences, their visions of self, and their visions of particular features of the social world in order to gain some purchase on their "common-sense" understandings about these matters.

This memo begins with some initial remarks about cultural analysis in sociology via ethnographic interviews. It then turns to considerations of the ethnographic interview in terms of methodological rigor and issues concerning interdisciplinarity. I provide some general commentary on each point and then raise what I maintain to be a central standing issue pertaining to each.

Taking a Stab at Cultural Analysis Through Ethnographic Interviews

A guiding presupposition for employing ethnographic interviewing is that it allows for researchers to grasp how individuals make meanings of themselves and the social words that they inhabit. The basic research objectives for those who employ this methodological tool are to discern what people "know" about themselves and their social worlds, how that knowledge is socially constructed and disseminated, and how that knowledge affect the behaviors enacted by such people. Each of these points of consideration concern some aspect of the sociology of meaning-making. Even a cursory history of sociological investigations of meaning making is beyond the purview of this memo. Yet, it should suffice to state that much of the past three decades of sociological explorations of meaning making as a cultural process is an extension of the work of Clifford Geertz (1973). Geertz provided the social sciences with an argument that carved out a space for cultural analysis to move beyond studies of the values and norms circumscribing collective and individual action. More recently, a slew of sociologists have posited their own claims about the sociology of meaning-making – including Pierre Bourdieu (1990, Bourdieu and Wacquant 1992), Anthony Giddens (1979, 1984, 1987), Margaret S. Archer (1996, 2000, 2003), Robert Wuthnow (1987), among others -- such that this area has become a vast and rich sub-field of the sociology of culture.

As it pertains to my work, I regard ethnographic interviewing as a tool that helps broaden the parameters for cultural analyses of the urban poor. Throughout the history of American sociology, urban ethnography has provided vivid portrayals of these individuals. In doing so, ethnographers have sustained the notion that what people do conveys much about how they make meaning of themselves, other people, and varied aspects of the social world. This often is the case. However, a flaw in this approach is that analyses of behavior do not provide transparent reflections of individuals' underlying

thoughts. For example, the fact that an individual is chronically unemployed and does not go to work on a regular basis (their behavior), tells us little about the complexity of their thoughts about the design and structure of labor markets, and what he or she believes is their optimal fit in them (their thoughts). Similarly, if someone has opted not to continue high school this does not mean that he or she does not have a keen sense of how and why a college education might help one to get ahead in life. Thus, rather than simply looking at behavior, it is important to pay attention to what people articulate as their own understanding of how social processes work and how they might negotiate social terrain. I maintain that an absence of sufficient dialogue between scholars in urban poverty research (as a branch of the sociology of race and ethnic relations) and those in the sociology of culture has fostered this analytical shortsightedness. Consequently, I regard the ethnographic interview as an essential means of bringing into the cultural analysis of the urban poor those issues, perspectives, and concerns that sociologists of meaning-making have begun pursuing in the past two decades.

Rigor and Ethnographic Interviewing

As it pertains to the ethnographic interview, rigor largely centers on the extent to which rich and provocative accounts are drawn from research participants, the extent to which these accounts collectively constitute some clear position or vantage point concerning the research question at hand, and whether ambiguities and contradictions in the testimonies of such research participants are effectively documented and investigated.

This process of documentation and investigation involves something different from efforts to resolve or erase ambiguities and contradictions in the course of data analysis. Instead, the challenge at hand is to determine as best as the researcher can how and why contradictions and apparent ambiguities emerge in the testimonies of research participants. In doing so, the researcher remain committed to the notion that a complex and fluid social world sometimes demands that people contradict themselves (or otherwise sometimes remain ambiguous in expressing their views) while explaining who they believe themselves to be, what purposes their actions have, or how they think that the social world functions. Rigorous analyses of interview data, then, involve not only tabulating the styles, patterns, forms, or types of content of people's expressions, but also when certain expressions do not fit easily with others, and why that may be the case for the individual who has been asked to account for him or herself with respect to a set of issues or concerns.

In the course of conducting sociological research via ethnographic interviewing a core standard of rigor is to achieve a sense of empirical saturation, or to reach a moment in data collection or analysis when similar or consistent patterns of argument or articulation are being offered by the research participants. Such saturation presumably indicates that the meanings that are constructed, adopted, or employed by such research participants (who, themselves, reflect the larger social group or category that is the target of the researcher's investigation) reflect some durable, and purposeful, shared cultural artifact.

There is, however, a standing dilemma in trying to ascertain whether one has achieved saturation. That is because in order for a researcher to feel saturated by consistent patterns of expression that researcher must translate those expressions into schemata or frameworks that are a part of his or her own cultural interpretive repertoire. Hence, what may be consistent or coherent for research participants, given their own cultural frames of understanding, may appear to the researcher to be contradictory, ambiguous, or incoherent.

Some form of serial exchange between researchers and respondents must occur for these potential cultural divides to be recognized and reconciled, but rarely is their sufficient time and financial support for such follow up interaction (and this is aside from the issue of the political implications of suspending or advancing one kind of interpretive frame (for instance, the research participants') rather than another (such as the researcher's). This points to "the politics of interpretation" (as opposed to "the politics of data collection in intimate settings") as a crucial point for consideration in ethnographic interviewing. Furthermore, the extent to which a researcher may be differ from a research participant in terms of class, race or ethnicity, age, or any number of other demographic traits may (but not always, as I explore a bit later) heighten this concern.

I now turn to discussion of one specific matter, that of silence produced in the course of interviewing, that is often taken to stand in the way of data rigorous analysis, but, I believe, should be taken as an avenue toward it. Silence has emerged in the past decade as an important topic of concern in sociological research (Denzin 1997, Hertz 1997, Kvale 1996, Poland and Pederson 1998). Much of the emphasis on silence, however, explores its significance in one of three traditional forms. The first concerns how people intentionally employ it to protect themselves by not revealing too much to an investigator. The second concerns how silence results from cultural differences between the investigator and the people that he or she is investigating (i.e., the absence of lengthy discussion on a topic because the respondent does not understand how or why the investigator has framed an inquiry). The third, which is closely related to the second, is the emergence of silence as a result of being asked about a common, regular, nondescript aspect of one's life such that the respondent has difficulty providing an extended commentary on the matter (i.e., why someone starts eating a meal with the vegetables rather than the meat, or why a particular shirt or pair of socks was worn on a given day). The form of silence has received less attention from research methodologists is categorical silence. This concerns the lack of requisite understanding or insight by which to frame an elaborate response or commentary (Poland and Pederson 1998). The researcher's awareness of this form of silence comes from knowledge of possible alternative articulations or the existence of actors who, by virtue of certain circumstances or conditions, are able to provide more elaborate commentary than are other actors.

The analytical challenge confronting ethnographic interviewers is to develop logics for better addressing not only the "unsaid" in the commentary of research participants, but also the "less elaborated upon," especially if other respondents provide more lengthy or detailed responses to the same questions. This move involves a shift from comparing the content of texts to comparisons of the presence and absence of certain kids of commentary across texts. More importantly, this shift in emphasis involves developing research designs and protocols that privilege silence as a finding to be analyzed in the context of where other research participants provide voice. An enrichment of the tradition of ethnographic interviewing, then, involves not simply emphasizing mechanisms and techniques that may generate more commentary from research respondents, but critically attending to the analytical significance of the absence of commentary (the latter case often viewed as nothing more than a methodological flaw).

Reaching Across Disciplines

Prior to a serious consideration of how such standards of rigor may be applied across disciplines it is important to grasp more fully the ethnographic interview is employed and regarded in various disciplines. I do not know much about this. I do believe, however, that certain social sciences may embrace contrasting perspectives on the meaningfulness of contradictory or ambiguous commentary that is generated through the ethnographic interview. A significant question concerning disciplinary differences is whether researchers in certain disciplines believe that the research endeavor is complete

only if contradictions or ambiguities s are somehow resolved through serial interviews, intensive probing, or even possibly the deletion of a research participant's commentary all together in order to bring more order to the data, or to create a more transparent, and seemingly more lucid, data set. If this is the case, I believe that the social sciences would benefit greatly by cross-disciplinary conversations about the approaches taken toward handling ambiguous or contradictory commentary produced through ethnographic interviewing

A rather specific matter concerning ethnographic interviewing that does transcend disciplinary boundaries pertains to the quest to garner intimacy with respondents such that they provide the kind of in-depth commentary that this is the ultimate goal of this research activity.[76] A common feature of qualitative research endeavors (ethnographic interviewing as well as any other close-to-the-people-under-study research approach) is the quest to establish intimate or sustained interaction with research subjects in presumably "natural" settings. In essence, the researcher steps into, and to varying degrees shares in, the everyday social worlds of the individuals under study.[77]

An initial underlying presumption in this debate was that researchers who share membership in the same social categories as their respondents (the most common being race, gender, and class) were best suited to uncover ideas, arguments, and opinions about issues and concerns related to those people or to those social categories (Merton 1972). A corollary presumption was that those researchers who do not share such membership either had to work especially hard to acquire the trust and confidence of respondents, or else accept that their scholarly analysis and interpretation may not reflect the veracity, depth, or subtlety that emerges from so-called "insider" research. In reacting to these presumptions qualitative field researchers strove to address whether and, if so, how greater ease, comfort, comprehension, and transparency could be established in the course of research, especially if such researchers occupied extreme outsider statuses. These efforts led field researchers to explore more critically the epistemological implications of either working to further their insider status or to confront the problems resulting from their outsider statuses (Andersen 1993, Baca Zinn 1979, De Andrade 2000, De Vault 1995, Ladner 1973, Naples 1996, Venkatesh 2002, Wilson 1974).

As most of these discussions centered on exploring the possibilities for increasing, maintaining, or reconciling with the difficulties of securing insider status, an implicit value was placed upon the insider position as the location that is most conducive for data collection.[78] The belief was that functioning from this position would enable the researcher to acquire the most meaningful, accurate, and honest data. Outsider positions were taken to be less constructive, if not all together detrimental, for conducting qualitative research. However, it is not always the case that occupying outsider positions necessarily inhibits a researcher from acquiring rich and insightful qualitative data. Hence, a reconsideration of outsider statuses is on order because, as I discuss below, the maintenance of rapport in the field can be threatened, if not altogether ruptured, by certain kinds of insider statuses.

Contemporarily, the insider-outsider debate has reached a point where the rigid dichotomization of insider and outsider positions has been called into question. It has been argued more recently that the biases and shortcomings associated with a researcher's occupation of an outsider status can sometimes be overcome or managed by the researcher's explicit acknowledgment of the existence of social distance or categorical dissimilarities between him or her and the individuals under study. Indeed, more thorough assessments of a researcher's distance or dissimilarity to the people under study, coupled with the researcher's declaration that no attempt was made to artificially or simplistically reduce or resolve these circumstances, have been woven into many of the contemporary qualitative studies that involve

76 1. The following part of this section is derived from a lengthier argument about the relationship of researchers to their research participants (Young 2004).

77 2. Scholars who engage this form of research are forced to continuously reflect upon and account for the depth and quality of their relationships to the individuals, situations, and conditions comprising their research agenda. In fact, in the past thirty years has been a period of rich dialogue about these matters (commonly referred to as the insider and outsider debate, see Andersen 1993, Baca Zinn 1979, DeVault 1995, Merton 1972, Naples 1996, Stanfield 1993, Wilson 1974).

78 3. Venkatesh is a notable exception in that he emphasizes how an outsider status can advance the process of data collection.

extending beyond racial and ethnic boundaries (Bourgeois 1996, Lamont 2002, Venkatesh 2000, Waters 1999). Moreover, it has become customary to include an appendix or a preface that illustrates exactly how the researcher engaged the field and established rapport with the individuals who were the central points of concern in the research. Indeed, in some cases attention was drawn to these issues throughout the body of the work itself. The point of this effort was to demonstrate to audiences how much researchers were aware of biases or shortcomings in their approach to field sites, which then conveyed a sense of legitimacy about their resulting work.[79]

The effort to more explicitly and provocatively explore how insider-outsider categories apply to researchers is a key advance over early claims that any extreme outsider status threatens the validity of the research. Accordingly, another key advance in research over the past three decades has been to open up considerations of how outsider status factors in the fieldwork experience, particularly in the development of ties to informants and the cultivation of respondents. One such consideration is the researcher's documentation of how one or more of their outsider characteristics become relevant points of reference in the fieldwork experience. For example, one scholarly commentary (Naples 1996) documents how respondents who initially took the researcher to be an insider member of the community sometimes self-defined as outsiders because they felt themselves to adhere to different cultural practices and ways of thinking in comparison to other community members. Although this work was not concerned with any outsider statuses maintained by the researcher, it provided an important analytical space for re-thinking whether and how outsiderness may be a constructive factor in data collection.

Another commentary focused on respondents' inability to immediately discern whether the researcher shared their ethnicity as the respondents (De Andrade 2000). That inability became a crucial factor for creating conversations in the field about this identity, which ultimately led to what the researcher found to be a rich and insightful pool of data for her project. The author interviewed people of her own nationality (Cape Verdian) about their ethnic consciousness. However, in most of the interviews the respondents made it clear that were not immediately sure that De Andrade was, herself, of that nationality. In analyzing her experiences, De Andrade offered that the insider position should never be taken as static and durable, and that such shifts can be a causal factor for producing useful data.

A third commentary (Reinharz 1997) asserts that one or more of a researcher's multiple selves (e.g., race, gender, or class status) may become relevant in the interactive dynamics of fieldwork. If these selves do not already appear at first sight in the course of fieldwork any of them are susceptible to becoming visible to respondents or field informants. More importantly, respondents and informants may react to any of these in ways that foster, hinder, or dramatically conversations with the researcher. Hence, respondents and informants may use these features and characteristics to determine whether to regard a researcher as more of an insider or outsider, and adjust their interaction with the researcher accordingly.

These and other investigations have led to the contemporary assertion is that there is no singular insider or outsider position that researchers occupy during the course of fieldwork, but rather myriad positions and statuses that can be viewed by respondents either as insider or outsider depending on the social circumstances or conditions affecting the research endeavor (De Andrade 2000, Jackson 2001, Naples 1996, Reinhartz 1997). Accordingly, the distinction between insider and outsider status should best be thought of as an analytical rather than experiential divide. Moreover, it has now been accepted that insider and outsider positions are fluid as they are continually restructured, retained, and abandoned during the course of interaction between researchers and respondents (De Andrade 2000, Naples 1996,

79 4. In some cases, scholars chose to explain their work in this tradition by affirming how it concerns people, issues, or circumstances that are intimately associated with or a part of the researcher's own life experiences. One example of such an effort is Patillo-McCoy (1999), an African American ethnographer who studied the social dynamics of a class sector of the African American community in which she also holds membership. Another is by a white American anthropologist, Carol Stack (1974) who studied the family dynamics of low-income African American mothers while she, herself, was a young mother.

Reinhartz 1997, Riessman 1987, Song and Parker 1995). These more recent commentators have demonstrated that insider status, though often crucial for the pursuit of certain research questions, is neither easily attainable nor consistently maintainable.

It is now understood in contemporary social science inquiry that outsider status not only cannot be fully erased in the course of research. Moreover, it is also understood that such a status also can play a constructive role in the production of data. Despite these advances, an enduring value is still placed upon the insider status as the privileged position from which to converse with respondents. By this I mean that researchers ultimately aim to increase their insiderness even if they know that they must contend with the various issues concerning outsiderness.[80] Consequently, there lacks a more critical exploration of how insider status may, in fact, actually inhibit conversation during specific moments in fieldwork.

I maintain that consideration of the less-recognized virtues of outsiderness is necessary so that more thought can be pursued on the question of how insider and outsider statuses function together in fieldwork, especially how they both operate as providers of possibilities and problems in the field. Indeed, different kinds of data can be acquired when the outsider position is taken as a legitimate point of entry into field work. After all, it may often be the case that a researcher wants to know something about intimate matters concerning race and ethnicity that a respondent chooses not to introduce into public commentary because the researcher is of the same race or ethnicity, thus making the insiderness of the researcher problematic. It may also be the case that respondents want to communicate something to a researcher that comes from, and thereby represents, a world far from that which is familiar to the respondent, thus making researcher outsiderness an advantage. Either objective can be advanced by researchers who are intensely considerate of where they think they stand between the ends of insiderness and outsiderness and how they think that respondents are locating them.[81] The goal in advancing the virtues of outsider statuses, however, is to do so while preserving what has long-ago been posited as virtuous about insiderness.

Emphases on the insider status have been associated with the obligation of researchers to acknowledge and assert the cultural complexity of the people whom they study, and the strive to capture and represent the voices of these people to the best extent possible given that the researcher is the ultimate creator of the statement being made about them. However, researchers also must develop a keen appreciate and preparedness to make use of what being on the outside can do to cultivate discussion rather than hinder it. A responsible approach toward managing the simultaneous effects of both statuses is to create, develop, and sustain conversations in the field in ways that enhance a researcher's sense of how others are thinking of him or her. The aim of the researcher must be to work toward maintaining the values and perspectives that are associated with insiderness while being conscientious about and appreciative of what being on the outside means for advancing conversations with people. This means researchers' using Reinhartz's notion of multiple selves in order to think about the capacity for these selves to be connected to a range of insider or outsider positionings. While researchers cannot be in full control of how they are located by the people whom they study, they can

---

80 . While Naples (1996) has effectively problematized the notion of the insider status as the most relevant position for data collection her commentary does not comprehensively explore outsider statuses that apply to the researcher. Her point about outsider status concerns how respondents identify with that position. Furthermore, while De Andrade argues that insider status is neither immediately presumed nor static in field work on racial and ethnic concerns, her commentary implicitly validates the vision of the insider perspective as the ideal for qualitative inquiry. She does so by affirming that working toward increased insider status is the ultimate goal, and she does so at the expense of exploring how that effort might be a hindrance for data collection. The challenge remains to better situate the outsider status as a powerful position by which to explore social phenomena such as racial consciousness and race relations.

81 . Indeed, some of the most insightful findings from ethnographies of black Americans that were conducted by non-African Americans were predicated on the authors' making explicit mention of their initial lack of understanding or profound curiosity about some aspect of African American culture (even if such explication was presented in an apologetic or discerning tone). Ulf Hannerz's *Soulside* (1969) is replete with numerous testimonies from the author about how unfamiliar or intrigued he was with certain events or phenomena unfolding around him as he studied a poor-to-working class African American street in 1960s-era Washington, D.C.. It remains that many classic and contemporary ethnographies of poor black Americans done by non-African Americans rarely dwell in great detail upon the revelations brought to the author by his or her outsider status.

In the case of my own work considerations of the virtues and drawbacks of insider status has led me to think about how one of my outsider statuses (affiliation with a university, for instance) allowed my research participants to see me as a conduit to a world far beyond their own. They understood that I was going to take their messages, after embedding them in a discursive style suited for academia, to audiences that they would probably never access by themselves. Many of these men made my outsider status as a university student instrumental for their purposes in talking to me. Accordingly, this provided one context whereby I was enabled to think of my outsiderness as a virtue rather than a hindrance.

think about the fieldwork experience as involving an amalgamation of insider and outsider positionings that come together to open up as well as restrict access to data. The challenge, then, is for researchers to strive to maintain a critical reflexivity about their simultaneous insider and outsider statuses as they work to negotiate the ever-shifting terrain of relating to respondents in field research. Such conscientiousness will result in serious thought about the fact that any field researcher is already indelibly grounded in a particular outsider status; that of being an individual that enters a social setting not simply to engage it like other participants might, but to analyze and document something about it for audiences often far removed from it.

Conclusion: Future Pursuits and Possibilities

As for future considerations, I believe that the ethnographic interview can shed even more light than it already has on how people think about social institutions that are presumed in American society to be relevant to their lives, but in certain ways are not perceived as such by research participants. By this I mean that future research can explore more critically what people understand to be the ways and means by which people get ahead in American society given the changes in labor market dynamics and the impact of technology. I believe this work can be extended to considerations of how people make sense of schooling as a mobility-enhancing experience (for instance, looking carefully at how people determine the worth or significance of certain kinds of schools given their objectives in life, or even certain employment sectors). In essence, intensive consideration of how people think about social institutions and processes that are far removed from their everyday lives can advance understandings of what people are prepared to do (or not do) on their behalf in order to improve their life prospects. Serious consideration of these and similar issues allows for a re-positioning of the potential conflicts in interpretation between research respondents and researcher's because the researcher's sense of a respondent's ambiguity, contradiction, or confusion may signal precisely how others (e.g., potential employers, social workers, or educators, in the case of poverty research) may interpret that participant. Rather than resolving or reducing ambiguities of conflicts, this example points to a circumstance where the very appearance of such developments may result in important research findings.

References

Andersen, Margaret L. (1993). "Studying Across Difference: Race, Class, and Gender in Qualitative Research." In *Race, Ethnicity, and Research Methods*, edited by J. Stanfield II and M.D. Rutledge. Newbury Park, CA: Sage Publishers.

Archer, Margaret S. (1996). *Culture and Agency: The Place of Culture in Social Theory*. Cambridge; New York: Cambridge University Press.

Archer, Margaret S. (2000). *Being Human: The Problem of Agency*. Cambridge; New York: Cambridge University Press.

Archer, Margaret S. (2003). *Structure, Agency, and the Internal Conversation*. New York: Cambridge University Press.

Baca Zinn, Maxine. (1979). "Field Research in Minority Communities." *Social Problems,* 27: 209-219.

Bourdieu, Pierre. (1990). *The Logic of Practice*. Translated by R. Nice. Stanford, CA: Stanford University Press

Bourdieu, Pierre, and Loic Wacquant. (1992). *An Invitation to Reflexive Sociology*. Chicago: University of Chicago Press.

Bourgeois, Philippe. (1996). *In Search of Respect: Selling Crack in El Barrio*. New York: Cambridge University Press.

De Andrade, Lelia L. (2000). "Negotiations from the Inside: Constructing Racial and Ethnic Identity in Qualitative Research." *Journal of Contemporary Ethnography*, 29: 268-290.

Denzin, Norman K. (1997). *Interpretive Ethnography: Ethnographic Practices for the 21st Century.* Thousand Oaks, CA: Sage Publications.

DeVault, Marjorie. (1995). " Ethnicity and Expertise: Racial Ethnic Knowledge in Sociological Research." *Gender and Society*, 9: 612-631.

Geertz, Clifford. (1973). The Interpretation of Cultures. New York: Basic Books.

Giddens, Anthony. (1979). *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis.* London : Macmillan.

Giddens, Anthony. (1984). *The Constitution of Society: Outline of the Theory of Structuration.* Cambridge: Polity Press, 1984.

Giddens, Anthony. (1987). *Social Theory and Modern Sociology.* Cambridge: Polity Press.

Hannerz, Ulf. (1969). *Soulside: Inquiries in to Ghetto Culture and Community*. New York: Columbia University Press.

Hertz, Rosanna, ed. (1997). *Reflexivity and Voice*. Thousand Oaks, CA: Sage Publications.

Jackson, John L. (2001). *Harlem World: Doing Race and Class in Contemporary America*. Chicago: University of Chicago Press.

Kvale, Steiner. (1996). *Interviews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, CA: Sage Publications.

Ladner, Joyce. (1973). *The Death of White Sociology*. New York: Vintage Books.

Lamont, Michèle. (2002). *The Dignity of Working Men*. Cambridge, MA: Harvard University Press.

Merton, Robert. (1972). "Insiders and Outsiders: A Chapter in the Sociology of Knowledge." *American Journal of Sociology*, 78: 9-47.

Naples, Nancy A. (1996). "A Feminist Revisiting of the Insider/Outsider Debate: The 'Outsider Phenomenon' in Rural Iowa." *Qualitative Sociology*, 19: 83-106.

Patillo-McCoy, Mary. (1999). *Black Picket Fences.* Chicago: University of Chicago Press.

Poland, Blake and Ann Pederson. (1998). "Reading Between the Lines: Interpreting Silences in Qualitative Research." *Qualitative Inquiry,* 4: 293-312.

Reinharz, Shulamit. (1997). "Who Am I? The Need for a Variety of Selves in the Field." *Reflexivity and Voice*. Thousand Oaks, CA: Sage Publications.

Riessman, Catherine Kohler. (1987). "When Gender is not Enough: Women Interviewing Women." *Gender and Society,* 1: 172-207.

Song, Miri and David Parker. (1995). "Commmonality, Difference, and the Dynamics of Disclosure in In-Depth Interviewing." *Sociology,* 29: 241-256.

Stack, Carol. (1974). *All Our Kin: Strategies for Survival in a Black Community*. New York: Harper Colophon Books.

Stanfield, John H. and Rutledge M. Dennis (eds.). (1993). *Race, Ethnicity, and Research Methods*. Newbury Park, CA: Sage Publishers.

Venkatesh, Sudhir. (2002). "Doing the Hustle: Constructing the Ethnographer in the American Ghetto." *Ethnography*, 3: 91-111.

Venkatesh, Sudhir. (2000). *American Project: The Rise and Fall of a Modern Ghetto*. Cambridge, MA: Harvard University Press.

Waters, Mary C. (1999). *Black Identities: West Indian Immigrant Dreams and American Realities*. Cambridge, MA and New York: Harvard University Press and Russell Sage Foundation

Wilson, William J. (1974). "The New Black Sociology: Reflections on the 'Insiders' and 'Outsiders' Controversy." In *Black Sociologists: Historical and Contemporary Perspectives*, edited by J. E. Blackwell and M. Janowitz, Chicago: University of Chicago Press.

Wuthnow, Robert. (1987). *Meaning and Moral Order: Explorations in Cultural Analysis*. Berkeley: University of California Press

Young, Alford A. Jr. (2004). "Experiences in Ethnographic Interviewing about Race: The Inside and Outside of It." In *Researching Race and Racism*, edited by M. Bulmer and J. Solomos, London: Routledge.