

The Promising Integration of Qualitative Methods and Field Experiments

By
ELIZABETH LEVY PALUCK

Randomized field experiments should take a more central place in qualitative research. Although field experimentation is often considered a quantitative enterprise, this paper illustrates the compatibility of field experimentation with various types of qualitative measurement tools and research questions. Integrating qualitative and quantitative data within field experiments allows investigators to move past simple average treatment effects and explore mechanisms of the identified causal effect. A more novel proposal is to use field experimentation as the organizing methodological framework for archival, ethnographic, or interpretive work, and to use ethnographic methods as the primary source of measurement in “experimental ethnography.” Sustained research and theoretical specificity can address some of the seemingly incompatible features of qualitative and field experimental methods. For example, small sample sizes are acceptable as part of a research program, and some theories of historical patterns or rare events could be disaggregated into smaller cause-and-effect linkages to test with field experiments in theoretically relevant contexts.

Keywords: field experimentation; qualitative methods; experimental ethnography; interpretive methods; archival methods

Over the past few decades, a productive exchange in political science has explored the idea that qualitative research should be guided by the logic of mainstream quantitative and experimental methods (e.g., Brady and Collier 2004; Gerring and McDermott 2007; King, Keohane, and Verba 1994). Most of these

Elizabeth Levy Paluck is an assistant professor of psychology and public and international affairs at Princeton University. She is an affiliate of the Households in Conflict Network and the Network on Experiments in Governance and Politics. Her research examines the influence of media on tolerance, dissent, and civic participation in post-conflict societies, including Rwanda, the Democratic Republic of Congo, and Sudan. She is a recipient of a Henry Frank Guggenheim Foundation grant.

NOTE: The author gratefully acknowledges Lee Ann Fuji, John Gerring, Jessica Greenberg, and Christopher Muller for helpful conversations and feedback.

DOI: 10.1177/0002716209351510

discussions focus on the logic of regression for drawing inferences from observational data, setting aside experimentation as an ideal but rare path to causal inference. A perhaps unintended message of this discussion seems to be that experimentation is a method unrealistic for most qualitative research projects, and consequently that experimentation is more naturally a quantitative enterprise. In short, qualitative researchers can aspire to use experimental logic for constructing counterfactuals and drawing causal inferences, but cannot use actual experiments.

This essay contends that experimentation—specifically field experimentation—can and should be more central to qualitative research. The argument rests on claims about what field experimentation is as well as what it is not. Field experimentation *is* one of the strongest methods for inferring causal relationships in real-world settings; it is *not* inherently quantitative.

By randomly assigning units (individuals, communities, organizations) into two groups, field experiments can infer that differences between the groups are due to an intervening “treatment” (a media program, a land redistribution policy, an elite negotiation meeting) applied to one group and not to the other. The key advantage of field experiments is that they draw causal inferences without invoking the untestable assumptions about the groups’ *ex ante* comparability that plague observational research.

The most straightforward reason why field experiments are perceived as quantitative enterprises may be found in the psychological concept of the availability heuristic (Kahneman, Slovic, and Tversky 1982). Put simply, there are few available exemplars of field experiments incorporating qualitative methods or testing questions traditionally associated with qualitative investigations. Therefore, experiments are thought of as quantitative in nature. I argue that the lack of field experiments using qualitative measurement or addressing traditionally qualitative questions stems more from our inability to think outside of this heuristic than from unassailable methodological and epistemological divides.

Consider the potential of qualitatively oriented field experimentation using a recent outstanding set of field experiments on gender and political leadership. Chattopadhyay, Duflo, and colleagues (Chattopadhyay and Duflo 2004; Beaman et al. forthcoming) tested the effect of having women in leadership positions on public expenditures and on the gender attitudes and political attitudes and behavior of constituents. The investigators capitalized on a policy experiment in India in which the government randomly reserved the position of village council leader for women candidates in one-third of all village councils in West Bengal. The investigators collected primarily quantitative evidence from voters and from records of public expenditures both in villages that did and did not reserve the leadership position for a woman.

The authors uncovered hugely consequential results. In some cases, women leaders increase women’s political participation. Women leaders consistently distribute public goods differently than male leaders according to their own preferences rather than female constituents’ complaints, and weaken stereotypes about women’s place in the public sphere. However, only after long-term exposure does voter approval of women’s leadership rise.

The point to take from this example is not that qualitative measurement would have made the experimental results “richer” or more detailed, although that is certainly the case. Using qualitative research methods in this field experiment could have provided a different understanding of the causal effect, identified possible causal mechanisms of change or sources of variation in the average treatment effect, and framed new interpretive understandings of authority, democracy, and gender within an experimentally assessed instance of social change.

For example, participant observation of women leaders *outside* of the council settings—such as in their homes, where they visit with other women—could have revealed whether they were influenced by women constituents in these more informal settings. Intensive interviews could compare social processes in villages with female or male council leaders to reveal how beliefs about women leaders’ efficacy shift. For example, did other council members, elders, or religious leaders make public statements about female leaders or the reservation system? Was there a tipping point at which common sentiment in villages with female leaders diverged from villages with male leaders? Such qualitatively generated insights could have enabled this study to contribute more to general theories of identity, leadership, and political and social change. Moreover, ethnographic work could compare understandings of authority and political legitimacy in villages with female- and male-led councils. Do the first female leaders inspire novel understandings of female authority and legitimacy, or are traditional gender narratives invoked just as frequently to explain women’s new power and position? Qualitative methods are uniquely positioned to answer these questions.

This brief example also makes apparent potential problems in integrating qualitative methods and field experimentation. For example, many qualitative methods involve a greater time investment and more fieldwork than quantitative data collection; is the extra time feasible or worthwhile in the context of a field experiment? How could participatory or ethnographic methods measure outcomes and processes in a sufficiently large enough sample for experimental comparison? More challenging, how can field experiments help investigate traditionally qualitative or observational research questions about historical patterns, institutions, elites, or rare events?

In the rest of this article, I address these concerns and expand upon some ideas about the integration of field experimentation and qualitative methods and questions. I first describe the benefits of integrating qualitative and quantitative measurement within a field experiment, using concrete examples from my own experimental work in Central Africa. I argue that qualitative measurement within a field experiment leads to a better understanding of the causal effect, suggests plausible causal explanations, and uncovers new processes that are invisible from a distance. Next I turn to more intensive tools of qualitative inquiry, such as ethnography and interpretive work. These methods can magnify cases, social processes, and concepts within an experiment, and in some cases provide the primary data for causal inference in what Sherman and Strang (2004) term “experimental ethnography.” I explore concerns about small sample sizes and scarcity of available units for random assignment. Finally, I turn to questions traditionally addressed

by qualitative and observational research, including questions about historical or rare events. I propose that field experiments have a role to play in many of these questions, which would require disaggregating complex theories and using theory to specify a universe of cases for present-day experimental tests.

I address this article not only to qualitative researchers, as encouragement to consider the use of field experimental methods, but also to field experimentalists, as inspiration to adopt more qualitative approaches in their research.

Integrating Quantitative and Qualitative Measurement within Field Experiments: More than the Sum of Its Parts

Collecting numerical, categorical, and ordinal data simplifies comparisons between experimental groups, but researchers could just as well collect and compare qualitative data from interviews, participant observation, and archives. It is widely recognized that inference is best supported by an integration of both types of data. Qualitative data can strengthen, modify, or altogether change the interpretation of quantitative data and describe important contemporaneous conditions of change.

Integrating quantitative evidence with qualitative evidence is especially appropriate for field experimental research, which, relative to laboratory experimentation, captures behavior in complex real-world settings. Field experiments typically measure the outcomes of behavior, but seldom capture what that behavior in fact consists of. Qualitative methods of investigation are best equipped to explore the meanings of the behavior in the context of the study, possible social and political dynamics by which the behavior is produced, ripple effects, and so forth.

A field experiment I conducted in eastern Democratic Republic of Congo (DRC) randomly assigned half of the region's radio antennae to broadcast a talk show. The talk show aired after a conflict reduction soap opera that was broadcast across the entire region (Paluck 2008b). The experiment asked if the talk show could increase face-to-face discussion about conflict reduction and if this discussion could produce more favorable attitudes toward community conflict reduction techniques. The outcome measurement, conducted in the randomly assigned regions with and without the talk show, included a close-ended survey instrument as well as a quantitative and qualitative behavioral measure.

In the behavioral measure at the conclusion of the survey, surveyors presented each study participant with a two-kilogram bag of salt. Surveyors told participants the salt was a thank-you gift for participating in the interview. They then added that a local NGO had identified a group in their community that was in need. They asked participants if they would like to donate any portion of their salt to this group. If participants asked, "Which group?" as all but four of them did, surveyors responded per a prewritten script: "Is there a group you would feel uncomfortable giving this to?" Nearly every participant responded by citing a disliked group:¹

“Yes, the (Banyamulenge/Rega/FDLR).” To this, surveyors responded, “Actually, that is the group for whom the donation drive is intended—would you still like to give?” As participants poured some amount of salt into a bag presented by the surveyor, or tied up their bag in preparation to store it away, they discussed their reasons for giving or not giving, their feelings about the donation, their expectations of the consequences of the donation, and their history with the least-liked group. The surveyors recorded this discussion as best they could by hand.

The strength of this mixed qualitative and quantitative measure was fourfold. I was able to record quantitative measures of whether and how much salt each participant gave (which I measured to the gram at the end of each day of interviews); qualitative data on the identity of each respondent’s disliked group; and data on participants’ reasoning, feelings, and expectations about helping or not helping this group. To measure the impact of the radio talk show, I used these data and the survey responses to compare listeners in the talk show broadcast regions to listeners in the non-talk show regions.

First I coded the qualitative discussions about the salt, which ranged from discussions about norms of sharing (“Congolese must pass on a gift”) to expressions of empathy and perspective taking (“When I am in great need, I know how much help from a stranger means to me”), to strategic reasons (“If I give them this salt, perhaps they will stop targeting my family”), to expressions of pure outrage (“They have killed family members, made us poor—I would rather die than help them”). This information would have been theoretically informative even if it had not taken place in an experimental context. It allowed me to explore reasons and motivations expressed by participants of varying economic situations, levels of education, experiences of the ongoing conflict, and other variables from the quantitative survey.

However, within the experimental context, these qualitative data significantly strengthened my interpretation of the radio show’s effect. Quantitative survey responses showed a *negative* impact of listening to the talk show—talk show listeners compared to listeners in non-talk show regions were less likely to endorse ideas in support of conflict reduction vis-à-vis their least-liked group and were more likely to endorse statements such as “violence is sometimes necessary in Congolese politics.” The salt measure showed that talk show listeners were also significantly *less* likely to donate their salt to the needy but disliked group (74 percent of control area listeners donated salt, while 55 percent of talk show area listeners donated). The qualitative discussions pointed in the same (negative) direction as the quantitative survey information regarding the impact of the talk show. I further discovered that radio listeners in the talk show areas expressed significantly more outrage and grievance against the least-liked group in their discussions about the salt (controlling for actual reported human rights abuses). The fact that these qualitative data were collected with a different instrument than the quantitative data strengthens the inference that encouraging discussion through a radio talk show had a negative impact on listeners. Even stronger triangulation would have included qualitative observations or interviews at another time or in another setting.

Causal Explanation Generation

These qualitative findings suggest a causal explanation of this negative result, specifically that talk show–inspired discussion made grievances more salient to listeners, reminding them of the hurtful actions of the other side. In general, field experimental results become considerably more useful with these kinds of potential explanations for the process or mechanisms of change. Theory can direct a researcher's eye toward particular situations and data sources that may explain the causal chain of events, but for more exploratory research (e.g., the effect of childhood abduction into a militia on political participation as an adult; Blattman 2009), deep contextual absorption (“soaking and poking” in qualitative lexicon; Fenno 1986) can inductively suggest explanations of experimentally assessed effects.

In the example of child soldiering, Blattman uses semistructured interviews with former abducted child soldiers, community leaders, and social workers to explore explanations for the finding that former child soldiers are *more* likely to vote. The qualitative data suggest that experience in the militia endowed former child soldiers with a sense of leadership and with a higher degree of maturity (Blattman 2009, 243). Causal explanations suggested by such qualitative research can then be tested in successive field experiments. In my own research, I conducted the field experiment in eastern DRC because qualitative research in a previous field experiment testing conflict-reduction radio soap operas suggested that discussion was an important mechanism of the observed changes in social norms and behaviors (Paluck 2009). In this previous experiment, I collected systematic observations of groups listening to the treatment and comparison radio programs and found that listeners kept up a steady rate of interjections, commentary, and side conversations during the broadcast regarding plot developments and characters' behavior. Moreover, listeners lingered after the broadcast was over to share their reactions and digest the messages of the show with one another. I hypothesized that face-to-face discussion about media with community members would shape perceptions of socially acceptable behavior, at least in the confines of that group. The experiment in the DRC attempted to test this causal explanation with an experiment that randomly assigned encouragement to discuss a media program via a talk show.²

In sum, the appeal of field experiments for qualitative researchers is that they offer the opportunity to generate strong causal inferences while “extracting new ideas at close range” (Collier 1999). I suspect that researchers who have long embraced the idea of mixed methodology will readily acknowledge this point. However, despite general enthusiasm for the idea, mixed methods have not been a common feature of field experiments.

Ethnography, Participant Observation, and Interpretive Work

More challenging than combining qualitative and quantitative data within a field experiment is integrating into an experiment qualitative methods that require

intensive time investment and field engagement, such as participant observation, intensive interviews, thick description, or ethnography. This broad group of methods is often employed in the service of interpretive goals—for example, complicating, historicizing, and enriching understandings of social science concepts like culture, democracy, or power (Wedeen 2002). In some cases, it may be useful to embed interpretive work in a larger field experimental test. Below I describe how these methods can also be used in a field experiment to investigate causal claims.

One straightforward way to integrate deep interviews, case studies, or ethnographies into an experiment is to select a reasonable number of observation units for close examination in the experimental and the control groups (see Tarrow [2004] on framing qualitative investigation within quantitative projects). Policy experiments have used this strategy—for example, the Moving to Opportunity experiment in American cities, which tested the effect of giving housing vouchers to low-income residents so that they could move into better neighborhoods (Turney et al. 2006). Sociologists and anthropologists working on this project conducted repeated intensive interviews with selected men and women who were randomly assigned to receive or wait for the voucher. The interviews explored quantitatively measured outcomes such as basic daily functioning and depression, phenomena that often require a fuller contextual understanding. In general, a feasible number of cases for intensive qualitative measurement within an experiment could be randomly selected from each experimental group to explore the contextual nature and heterogeneity of the experimentally assessed causal effects.

“Experimental Ethnography”

A more ambitious proposal in this vein is to conduct ethnographic case studies for all of the units of observation in a field experiment in what Sherman and Strang (2004) term “experimental ethnography.”

Experimental ethnography is a tool for answering questions about why programmatic attempts to solve human problems produce what effects, on average, in the context of the strong internal validity of large-sample, randomized, controlled field experiments. . . . This strategy can achieve experiments that create both a strong “black box” test of cause and effect and a rich distillation of how those effects happened inside that black box, person by person, case by case, and story by story. (P. 205)

Writing from the perspective of program evaluators, Sherman and Strang discuss a recent randomly implemented policy for restorative justice in England and Australia. The policy invited victims, perpetrators, and all those affected by the crime to meet and discuss how the perpetrator should repay his or her debt to society. When police officers offered this program to untried perpetrators and their victims, they told each party that if both parties accepted, they would have a 50 percent chance of having the meeting because the program was in an experimental trial.

Sherman and Strang describe how ethnographies describing the experiences of victims, perpetrators, and their families during and after the restorative justice process would have been important for fully understanding the effects of this program.³ Specifically, experimental ethnography could use an iterative process of theory development and testing commonly associated with qualitative approaches, or grounded theory (Glaser and Strauss 1967). “The hypotheses that are generated from interviews or observations of one case can immediately be tested against new data on the same hypotheses collected on other cases. Even if these hypotheses and their tests are later reduced to quantitative form, the fact that they would not have emerged without ethnographic work provides a strong justification for the added cost and effort of experimental ethnography” (Sherman and Strang 2004, 211).

Qualitative data on the victim’s reaction to the crime, in Sherman and Strang’s example, suggested the hypothesis that the magnitude of potential benefit of restorative justice on the victim’s mental health was directly proportionate to the magnitude of the harm the victim suffered from the crime. The qualitative evidence both “discovered” this grounded claim and offered a way to test it, through continuous comparisons between treatment and control groups. This example also illustrates the ability of qualitative data to discover interactions, or systematically different responses to the experimental intervention. Discovering interactions pushes our understanding of the causal effect beyond a more simplistic average treatment effect. Sherman and Strang note that it is best to conduct this kind of theory testing when *all* of the cases in an experiment can be included in an ethnography, which should be feasible for “samples of a hundred or so” (2004, 211).

Small N Concerns

The Sherman and Strang proposal exposes an important trade-off, the classic tug of war between breadth and depth that typically leaves qualitative researchers with a small sample size. Other times, qualitative researchers’ sample size is restricted by the limited number of units to study—for example, only six countries that meet the criteria for a certain research topic, or 12 non-overlapping broadcast areas in the region of interest. I have two suggestions regarding this trade-off.

Collaboration is one answer to the problem of conducting ethnography with all of the units of an experiment. Several qualitative researchers working as a team could each take responsibility for a portion of units in the treatment and control groups. Researchers’ responsibilities should overlap for a few units, as the overlapping ethnographies could serve as a continuous check on the comparability of their methods and observations. This kind of collaborative ethnography, which is not uncommon in sociology (Wilson and Taub 2006), has the potential to provoke a productive discussion among ethnographers regarding the comparative versus particularistic nature of their work. The challenge of comparing their ethnographic data in the service of drawing causal inferences

would require ethnographers (or interpretivists, participant observation researchers, and the like) to make their process—their definitional terms, observational procedures, and selection of place and subjects—more transparent and replicable. Such an effort only would succeed by increasing the comparative nature of the ethnographic enterprise. While some ethnographic traditions (particularly in anthropology) are opposed to the idea of producing replicable procedures and observations, this kind of a collaborative work would advance the comparative goals of ethnographers who are amenable to the idea.

A sustained research program is another way to accommodate a small sample in a field experiment.⁴ With a small sample size, researchers may not be able to identify modest or small effects, or may overestimate or underestimate larger effects. My collaborator Donald Green and I have argued that in this case it is still worthwhile to do the experiment in the context of a sustained research program (Paluck and Green forthcoming). Repeated experiments on the same general question will average out to the true unbiased effect, and the cumulative sample size will improve the precision of the estimated effects.

Treating Questions Typically Associated with Observational and Qualitative Investigation

One of the most frequently voiced reasons for not using field experimental methods is that a certain class of research topics is too historically based or would be unethical or impossible to test using random assignment. Questions about the historical pattern of state formation, the causes of revolutions or genocides, elite decision-making about nuclear deterrence, and the democratic peace hypothesis all fall into this category. These topics are sometimes cited as evidence that observational and qualitative researchers struggle with more “important” or “bigger” questions than those addressed by experimental methods.

Of course, field experiments (and as-if-random “natural” experiments; Dunning 2008) have already addressed many important questions that seemed unsuited to experimentation prior to their successful execution. To date, and mostly without the explicit use of qualitative methods, experiments have answered questions about the effects of political campaigns (D. Green and Gerber 2008; Nickerson 2008; Wantchekon 2003), police raids and crime deterrence (Sherman et al. 2002), mass media programming (Paluck 2009; Green and Vavreck 2008), ethnic diversity (Habyarimana et al. 2007; Posner 2004), international election monitoring (Hyde 2007), deliberative democracy techniques (Farrar et al., forthcoming; Wantchekon 2008), gender and politics (Beaman et al. forthcoming; J. Green 2008), corruption (Olken 2007), employment discrimination (Pager 2007), educational attainment (Sondheimer and Green forthcoming), healthcare (King et al. 2007), slavery and trust (Nunn and Wantchekon 2009), and child soldiering (Blattman and Annan forthcoming). Thus far, I have argued that including qualitative methods can extend the reach of field experimentalism further.

Still, causal questions rooted in history or addressing elites, violence, and country-level and rare events like social movements and revolution are at one level beyond the reach of experiments. Random assignment of the purported causes of these events would be unethical or logistically impossible without dictatorial powers or a time machine. One point made in response to this dilemma is that relatively more narrow field experiments accumulate the “stubborn facts that inspire theoretical innovation” (D. Green 2005, 100). Field experiments gradually collect unbiased causal facts upon which a more complex theory can be built.

I propose another idea that flows in the opposite direction. In contrast to building theories from relatively narrow empirical facts, investigators could start at the level of their highly complex theories and disaggregate them in a way that would make field experimentation possible for a few of the causal links in their specified chain. Theories of genocide, for example, make many causal claims about the road to violence. Some purported causes of genocide include elites threatened by a shift in power, bureaucratic or other tools for ethnic differentiation, land shortages, and so forth. A field experiment could not and would not randomly assign all of these conditions, but it could, for example, examine the effects of policies (introduced progressively in randomly assigned areas of the country or subsets of the population, i.e., a “random rollout”) that increase or decrease ethnic differentiation (identity cards or citizenship papers) or redistribute land.

Integrating field experiments into these traditionally observational research programs in this manner would require theoretical specificity, strategic case selection (for which qualitative researchers are exceptionally qualified), and (in some cases) cooperation with policymakers or political elites. Researchers would need a high degree of theoretical specificity and clarity in their definitions of the necessary contextual conditions of a present-day theoretical test. Some theories are intended only for historical cases (Skocpol 1979); in these instances, field experimentation is obviously impossible. But for theories intended to extend into present-day contexts, researchers would need to draw out sufficient and necessary conditions for the field experimental context.

Using theories that describe necessary and sufficient conditions of the phenomenon of interest, qualitative researchers have honed the skill of case selection (Seawright and Gerring 2008) into a systematic method that requires deep contextual and historical knowledge. Selecting present-day relevant cases would be the critical task for researchers testing theories of historical events with field experiments. Finally, many such field experimental tests would probably require collaboration with policymakers and political elites, since many of these kinds of questions involve structural, economic, or institutional shifts. Many relevant changes are occurring through new policies (recall land policies in developing countries), which could be rolled out randomly. Collaborating with governments and nongovernmental and/or international organizations presents a host of ethical and practical dilemmas, but it should not be written off as impossible. Currently, field experimentation is receiving increased respect and interest from policymakers and international organizations, mostly on the wings of the influential

movement to include field experiments in development economic policy and from efforts of some political scientists as well.⁵ As economists have proved with the development community, a few very useful experiments can interest stakeholders in fielding and participating in experiments of their own. Experimentation with (and even on) political elites would make the use of experimentation in observational research programs more of a possibility.

Conclusion

Researchers should not foreclose the possibility of using field experiments in qualitative or observational research programs or using qualitative measurement in field experiments. Integrating field experimentation into a qualitative research program will be a difficult but creative and productive process. It will require knowledge of the cases, theoretical clarity, and comparable and meaningful outcome measures. Qualitative methods, from case selection to interviewing to participatory observation, are all necessary on some level to conduct good field experiments. For this reason, qualitative researchers and current field experimentalists alike could benefit from collaboration. Integrating qualitative methods with field experiments should encourage new and interesting investigator partnerships and learning within all types of methodological persuasions.

Notes

1. Based on the “content-controlled” technique pioneered by Sullivan, Piereson, and Marcus (1982).
2. Note that in the DRC experiment, I was missing a critical arm of the experiment (due to logistical reasons) in which a third control group did not have access to the soap opera, which provided the topics of discussion, or to the talk show, which encouraged the discussion. Including a no-soap, no-talk control group would show (1) the effect of the soap opera and (2) the additional effect of discussion inspired by the talk show about the soap opera. I am implementing this design in a new experiment on a peace and democracy radio campaign in Southern Sudan, by randomly assigning a radio show, discussion, radio show plus discussion, or no intervention (Paluck 2008a).
3. They also suggest that ethnographies of the victims and perpetrators who did not accept the offer to be a part of the program would have helped explore the reach of the restorative justice program, and also more generally the ability of experimental trials to measure causal effects in a *representative* portion of the population.
4. Besides the problem of low power to detect causal relationships, small samples mean that simple random assignment is more likely to create an unbalanced comparison. For example, in a sample of twelve manufacturing companies, a random “run” of similar assignment numbers could assign all five car companies in the sample to the treatment condition. This problem of balance can be addressed by matching procedures prior to randomization—simple stratification procedures in which randomization is conducted within stratified groups of car and drug manufacturing companies, for example, or more complex matching with multiple strata using statistical software (e.g., Coarse Exact Matching; Iacus, King, and Porro 2008). In my small N experiments in Central Africa, I have randomized within stratified villages and broadcasting regions.
5. EGAP, or Experiments in Governance and Politics, is one example of a recent organizational effort involving political scientists and policy organizations.

References

- Beaman, Lori, Raghabendra Chattopadhyay, Rohini Pande, and Petia Topalova. Forthcoming. Powerful women: Does exposure reduce bias? *Quarterly Journal of Economics*.
- Blattman, Christopher. 2009. From violence to voting: War and political participation in Uganda. *American Political Science Review*, 103 (2): 231-47.
- Blattman, Christopher, and Jeannie Annan. Forthcoming. The consequences of child soldiering. *Review of Economics and Statistics*.
- Brady, Henry, and David Collier, eds. 2004. *Rethinking social inquiry: Diverse tools, shared standards*. Lanham, MD: Rowman & Littlefield.
- Chattopadhyay, Raghabendra, and Esther Duflo. 2004. Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica* 72 (5): 1409-43.
- Collier, David. 1999. Data, field work and extracting new ideas at close range. *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 10 (1): 1-2, 4-6.
- Dunning, Thad. 2008. Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly* 61 (2): 282-93.
- Farrar, Cynthia, James Fishkin, Donald P. Green, Christian List, Robert Luskin, and Elizabeth L. Paluck. Forthcoming. Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science*.
- Fenno, Richard F. 1986. Observation, context, and sequence in the study of politics. *American Political Science Review* 80 (1): 3-15.
- Gerring, John, and Rose McDermott. 2007. Experiments and observations: Towards a unified framework of research design. *American Journal of Political Science* 51:688-701.
- Glaser, Barney, and Anselm Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Green, Donald P. 2005. On evidence-based political science. *Dædalus* 134 (3): 96-100.
- Green, Donald P., and Alan S. Gerber. 2008. *Get out the vote: How to increase voter turnout*. 2nd ed. Washington, DC: Brookings Institution.
- Green, Donald P., and Lynn Vavreck. 2008. Analysis of cluster-randomized experiments: A comparison of alternative estimation approaches. *Political Analysis* 16 (2): 138-52.
- Green, Jennifer. 2008. Mobilizing women to vote in traditional societies: An experiment encouraging political participation in rural India. Working paper. Institution for Social and Policy Studies, Yale University, New Haven, CT.
- Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein. 2007. Why does ethnic diversity undermine public goods provision? *American Political Science Review* 101 (4): 709-25.
- Hyde, Susan. 2007. The observer effect in international politics: Evidence from a natural experiment. *World Politics* 60 (1): 37-63.
- Iacus, Stefano M., Gary King, and Giuseppe Porro. 2008. Matching for causal inference without balance checking. Harvard University Working Paper. <http://gking.harvard.edu/files/abs/cem-abs.shtml>.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas. 2007. A "politically robust" experimental design for public policy evaluation, with application to the Mexican Universal Health Insurance Program. *Journal of Policy Analysis and Management* 26 (3): 479-506.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Nickerson, David W. 2008. Is voting contagious? Evidence from two field experiments. *American Political Science Review* 102 (1): 49-57.
- Nunn, Nathan, and Leonard Wantchekon. 2009. The slave trade and the origins of mistrust in Africa. Working paper. Harvard University. http://www.economics.harvard.edu/faculty/nunn/files/Trust_v6.pdf.
- Olken, Benjamin. 2007. Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy* 115 (April): 200-49.

- Pager, Deval. 2007. The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The Annals of the American Academy of Political and Social Science* 609:104-33.
- Paluck, Elizabeth Levy. 2008a. A field experiment testing discussion and media in Southern Sudan. Unpublished manuscript.
- Paluck, Elizabeth Levy. 2008b. Is it better not to talk? A field experiment on talk radio and ethnic relations in eastern Democratic Republic of Congo. Working Paper, Harvard University, Cambridge, MA.
- Paluck, Elizabeth Levy. 2009. Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology* 96:574-87.
- Paluck, Elizabeth Levy, and Donald P. Green. Forthcoming. Deference, dissent, and dispute resolution: A field experiment on a mass media intervention in Rwanda. *American Political Science Review*.
- Posner, Daniel. 2004. The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review* 98 (4): 529-45.
- Seawright, Jason, and John Gerring. 2008. Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly* 61 (2): 294-308.
- Sherman, Lawrence, Dennis Rogan, Timothy Edwards, Rachel Whipple, Dennis Schreve, Daniel Witcher, William Trimble, Robert Velke, Mark Blumberg, Anne Beatty, and Carol Bridgeforth. 2002. Deterrent effects of police raids on crack houses: A randomized, controlled experiment. *Justice Quarterly* 12 (4): 755-81.
- Sherman, Lawrence, and Heather Strang. 2004. Experimental ethnography: The marriage of qualitative and quantitative research. *The Annals of the American Academy of Political and Social Science* 595:204-22.
- Skocpol, Theda. 1979. *States and social revolutions: A comparative analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- Sondheimer, Rachel M., and Donald Green. Forthcoming. The Brody Paradox revisited: Using experiments to estimate the effects of education on voter turnout. *American Journal of Political Science*.
- Sullivan, John, James Piereson, and George E. Marcus. 1982. *Political tolerance and American democracy*. Chicago: University of Chicago Press.
- Tarrow, Sidney. 2004. Bridging the quantitative-qualitative divide. In *Rethinking social inquiry: Diverse tools, shared standards*, ed. Henry E. Brady and David Collier, 171-80. Lanham, MD: Rowman & Littlefield.
- Turney, Kristin, Susan Clampet-Lundquist, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2006. Neighborhood effects on barriers to employment: Results from a randomized housing mobility experiment in Baltimore. Working Paper #511, Princeton University, Princeton, NJ.
- Wantchekon, Leonard. 2003. Clientelism and voting behavior: Evidence from a field experiment in Benin. *World Politics* 55:399-422.
- Wantchekon, Leonard. 2008. Expert information, public deliberation, and electoral support for "good" governance: Experimental evidence from Benin. Working Paper. New York University.
- Wedeen, Lisa. 2002. Conceptualizing culture: Possibilities for political science. *American Political Science Review* 96:713-28.
- Wilson, William Julius, and Richard Taub. 2006. *There goes the neighborhood: Racial, ethnic, and class tensions in four Chicago neighborhoods and their meaning for America*. New York: Knopf.