

Can field work be “scientific”?

It is hardly a secret in the social sciences that evidence based on field work – interviews with and observation of actual humans in their natural habitats – is taken less seriously than evidence based on survey data or experiments. Audiences at job talks or conferences, as well as reviewers for journal articles, frequently give more weight to large-N statistical analysis, particularly experiments, even when the bulk of a project involved open-ended interviews or participant-observation. Given the amount of time and energy that scholars (including graduate students writing their dissertations) devote to field work, this is a sorry state of affairs. How can field researchers make their work as persuasive as possible to colleagues inclined toward other methods?

Field work obviously serves many worthwhile purposes: improving language skills, learning cultural norms, obtaining datasets that are otherwise unknown or unavailable, creating professional networks, gaining access to policymakers, and so forth (Moehler 2005). Field work is often extremely helpful for identifying interesting puzzles (i.e., choosing topics). It is typically essential for validating measurements and accurately describing what is actually happening on the ground. And it is extremely useful in formulating or refining contentions that can be tested independently (“hypothesis generation”).¹ Such benefits alone might justify the field work that scholars in comparative politics normally conduct. But many projects in political science, anthropology, and sociology use evidence collected in the field to determine whether a particular claim was correct or incorrect (sometimes referred to as “hypothesis testing”); some researchers rely primarily or even exclusively on interviews and observations in reaching their conclusions about what did or did not actually happen in a particular case. Such projects face an uphill battle in increasingly methods-conscious disciplines.

In this essay, I focus on how field researchers can address some of the most obvious methodological challenges common to ethnography and thus make their work as persuasive as possible. I argue that these challenges are not primarily related to the growing focus on causal identification in political science and economics, nor to the fact that information collected from interviews or observations may be qualitative in nature. Rather, they have to do with the mechanics of field work and the resulting degree to which audiences can feel that the work has been done properly. In other words, they have to do with establishing that researchers have taken “the kind of methodological care and transparency that gives other scholars confidence” in their conclusions (Stevenson 2005: 15).

I argue that field researchers can never fully address some of these methodological challenges, given the impracticality of perfectly replicating ethnographic research and the normal absence of double-blind controls. As a result, evidence from field work will often meet with a certain amount of justifiable skepticism. Nevertheless, the methodological problems that commonly characterize field work can be substantially alleviated, and a number of the remedial measures I propose are “cheap” in terms of effort expended. By adopting as many of them as possible, researchers can substantially strengthen their claims and justify greater reliance on the evidence that they collect.

I begin with a discussion the larger social scientific enterprise to elucidate the nature of the methodological vulnerabilities of field research (which are often implicit or poorly articulated in critiques of field work). Next, I discuss potential ameliorative strategies, including ways in which researchers can

¹Some scholars in the humanities conduct field work for purely interpretivist purposes; this essay focuses exclusively on the social sciences, though some of the points might apply for historians.

signal “methodological care and transparency” and competence when they cannot directly demonstrate these features of their work. Finally, I present an example of a graduate student research project based on field work that was placed on more solid footing by adopting some basic, practical steps.

The (social) scientific method

The scientific method consists of cycles of theorizing and testing. Researchers develop potentially testable propositions (often referred to as hypotheses) based on the results of prior research, casual observation, or deductions from well-grounded assumptions. They then attempt to verify such propositions empirically. The results of their investigation in turn allow them (or other researchers) to refute, affirm, extend, or qualify their original propositions. In theory, this iterative process leads to the accumulation of knowledge. Social scientists attempt to apply this same methodology to human behavior, and it tends to produce greater knowledge over time. To take one example from political science, an investigation into party systems might begin with the observation that English-speaking countries generally have two-party systems; this observation could then lead – through repeated cycles of theoretical extrapolation and empirical verification – to a more refined causal contention that relates the degree of proportionality in the electoral system to the number of parties in each electoral district.

Each piece of this enterprise is premised on certain assumptions. “Hypothesis generation”, for instance, should favor propositions that are logically coherent, plausible, and empirically verifiable. (Hypotheses that are not logical cannot be correct, regardless of the empirical evidence that would seem to support them; hypotheses that are not *ex ante* plausible are not worth testing; and hypotheses that cannot be subjected to empirical verification fall outside the scope of the scientific enterprise.) For its part, “hypothesis testing” is guided by the notion of *ceteris paribus*: the attempt to isolate specific causal effects under certain scope conditions.

When it comes to hypothesis generation, field work is not necessarily disadvantaged vis-à-vis other methods: watching and talking to people is a perfectly reasonable way to develop ideas about why they do what they do. In fact, field work is probably more likely to generate plausible and empirically verifiable hypotheses than many other methods (e.g., deductions from some set of assumptions about human behavior that themselves may or may not be particularly plausible).

The challenge comes with hypothesis testing: evidence based on field work is often challenged on the grounds that researchers have failed to accurately describe what is happening (sometimes), to establish the direction of causality (occasionally), to rule out alternative explanations (often), or to establish the generality of their findings and properly specify their scope conditions (often). At a general level, almost all social science methods are subject to the same critiques mentioned above. For instance, studies based exclusively on the statistical analysis of observational data are even more vulnerable to the first three critiques than studies based on field work. And even randomized, controlled experiments in a laboratory setting – the gold standard for establishing causality and isolating particular effects – are frequently vulnerable to the last criticism. What really causes problems for field research concerns the mechanics of the work – that is, the way interviews are conducted and observations made. Doubts about this part of the scientific process affect the degree to which studies based on field work are taken seriously by other scholars and, in turn, can serve as the starting point for future work.

The professional presumption

For the scientific method described above to lead to knowledge accumulation, researchers must presume that other researchers have done their work honestly and competently. This “professional

presumption” is only *prima facie* rather than absolute – it may turn out that investigators have made errors or even fabricated their results – but without at least a weak prior that other research was done properly, individual studies cannot build into a larger scholarly edifice.

Not surprisingly, the scientific community has developed a number of norms and practices that encourage honesty and competence, thus lending credence to the “professional presumption”:

- Oversight structures (laboratories, professional associations, etc.) that prevent or sanction manifest carelessness and fabrication.
- Professional training and credentialing, which help researchers to avoid common errors, encourage them to internalize ethical norms, and weed out manifestly incompetent or mendacious researchers.
- Adherence to shared standards for adjudicating empirical claims, in order to ensure that researchers are interpreting the same results in the same way.
- Conventions for reporting the methods used in a study, so that other researchers can interrogate them and, if appropriate, replicate the study.
- Conventions regarding the sharing of data collected in the course of the study, to allow replication of specific analyses by other scholars (even if the original data collection itself is not replicated).
- Double-blind controls, which reduce the chances that researchers inadvertently bias their findings toward their own priors when conducting the study.

However, there is considerable divergence within the social sciences in the degree of fidelity to the norms listed above. And it is here that traditional field work often compares poorly to other methods.

Studies based on statistical analysis, for instance, usually meet most if not all of the criteria above. Researchers take classes – and sometimes qualifying exams – that attest to their competence in using particular tools and familiarize them with professional standards for using these tools (e.g., the need for tests of robustness). Such classes typically include discussions of common ethical pitfalls (e.g., data dredging and p-hacking). Researchers rely on the same standards for adjudicating hypotheses (e.g., $p < .05$) and – most of the time – for determining whether the data they collect is measured and recorded properly (e.g., statistical tests of inter-coder reliability). Analysis is often conducted on datasets that are ultimately available to other scholars, and replication exercises have become increasingly common in many disciplines; scholars now routinely save not only their data but also their coding files for that purpose. Finally, although scholars often both collect and analyze the same data in a study, thus destroying any pretense of double-blind controls, many researchers also work with statistical data that is not of their own confection; in such cases, the data collection piece of a project is arguably double-blind.

To be sure, fidelity to scientific norms is hardly perfect in statistical analyses. Scholars working with novel or proprietary data cannot prove up-front that they have reported only robust results, refrained from data-mining, and the like; they must implicitly request some degree of trust from the audience. Likewise, experimental research can be highly sensitive to the craft of the researcher, and double-blind controls are often absent in field experiments. But challenges in justifying the “professional presumption” tend to be more acute in field work, for several reasons.

First, many field researchers receive no technical training in how to conduct interviews or participant-observation. Fewer still have earned professional credentials that attest to their competence and cognizance of ethical pitfalls (e.g., how to handle discrepant reports from informants). Some – including many graduate students on the job market – simply have no experience conducting interviews. This

problem is particularly pronounced for students in economics, who may feel uncomfortable with the whole idea of talking to subjects (even when it is obviously the right course of action from a probative perspective). But, with the exception of anthropologists, even those who do a good deal of field work typically receive limited training in such techniques.

It is difficult to overstate the problem that inexperience and inadequate training can create. Poorly conducted interviews are like poorly crafted experiments: they simply fail to generate useful information. Researchers may simply have never asked the “right” question of their informants, or failed to ask it in the right way, and they never know what they missed. Although many recent field experiments in political science and economics are probably afflicted with similar problems, the same is generally not true for most laboratory experiments and statistical analyses.

Second, the way field researchers present their findings often makes it difficult to interrogate their method or reassure colleagues about their competence. Many field researchers do not make public their interview guides or systematically record the conditions under which interviews were conducted. Indeed, field researchers frequently do not spell out the sampling strategy for their informants (including which types of people declined to participate and how researchers identified replacements for them). Lack of transparency means that other scholars cannot evaluate what researchers have actually done (or left undone) and how well they did it. The degree to which such problems call into question the “professional presumption” is exacerbated if the researcher in question has had no formal training in his principal method of data collection.

Third, barriers to replication loom larger in field work. Data collection (i.e., doing the interviews) often cannot be replicated for one of many reasons: because informants’ identities are confidential; because they have died, moved, or forgotten things; or because conditions under which they were originally observed or interviewed have changed significantly. Indeed, pure replication of data collection may be impossible even in theory, because informants sometimes impart different information simply because they are being interviewed or observed for a second time.

Replication of data “analysis”—that is, reviewing the evidence to determine whether it systematically supports an author’s interpretation—is only possible when the raw data on which it is based are available. In the case of field work, such raw data typically includes notes from interviews and observations. There are often legitimate reasons for not making certain material available, such as confidentiality of sources. But much of the time, replication is made impossible simply because researchers find it inconvenient to make material from their interviews and observations available. One promising step forward comes from the guidelines for submission to the *American Political Science Review*, which requires the release of such materials.² However, such a degree of transparency is not yet standard practice in the discipline.

Fourth, lack of double-blind controls presents special problems for field researchers. In virtually all field work-intensive studies, the same researcher develops the central hypothesis, collects the information to

²“Where field or observational research is involved, anonymity of participants will always be respected; but the texts of interviews, group discussions, observers’ notes, etc., should be made available on the same basis (and subject to the same exceptions) as with quantitative data.” (<http://www.apsanet.org/apsrsubmissions>, accessed July 7, 2015)

test it, and analyzes the evidence collected. Indeed, field researchers cannot avoid knowing why they are observing a particular group or interviewing a subject. This fact obviously opens the door to a wide variety of biases. Again, a number of recent field experiments in political science and economics may also fail this test, but the problem is not inherent in their method.

Finally, it is unclear whether different researchers employ the same criteria for determining when they have proven their claim. It remains relatively rare for researchers to be explicit about what sorts of the evidence would be required to confirm or disconfirm a particular contention. Field researchers rarely specify which people would need to be interviewed for a case to be adjudicated in one way or another, or what questions would have to be asked of them. And even when researchers offer a compelling standard for treating a *case* as confirmatory or disconfirmatory, they rarely state explicitly how a specific observation, comment, or interview should be regarded. The result is a fuzzy standard for adjudicating among rival hypotheses (or between the central hypothesis of interest and the null).

Some of the problems mentioned above – e.g., potential ineptitude in conducting interviews as a result of inadequate training – may make it more difficult for a researcher to verify *any* proposition (i.e., bias toward the null). But most of the bias is likely to be confirmatory. In other words, it is entirely possible that scholars will conclude their favored hypothesis has been confirmed when the truth is otherwise. And audiences rarely have enough information about how the research was conducted to be reassured on this score.

When scholars cannot address the problems discussed above, their research takes the form of a whopping “trust me” – that is, an assertion that the interviews and observations were conducted competently, interpreted accurately, and summarized faithfully without any supporting documentation to that effect. This assertion, of course, is unlikely to convince skeptics. Although some social scientists might foolishly disregard or denigrate ethnographic work regardless of how well it was conducted, a degree of skepticism of conventional field work is indeed justified.

Skeptical reactions are likely to be particularly common when the researcher in question is junior, for several reasons. First, graduate students and junior scholars generally have less experience conducting field work and may thus be apt to make more errors. Second, younger scholars have generally not had the chance to acquire a personal reputation for honesty and competence on which other researchers can rely when it is impossible to directly verify whether a study was well-executed. Finally, junior researchers may have stronger incentives to interpret evidence as confirmatory in order to get a job or first publication (or at least are perceived as having such incentives). For these reasons, conventional field work often fails to persuade audiences at job talks and reviewers at top journals.

Improving field work

The challenges identified above cannot be fully resolved. For instance, the practical barriers to re-contacting prominent subjects are often insurmountable. Likewise, it would be extremely difficult to introduce full double-blind controls, because for many social scientists, the relationship between data collection in the field and “write-up” of their field notes is a matter of tradecraft.

That said, field researchers can do much more to ameliorate many of these challenges. With that more modest objective in mind, Table 1 below offers some suggestions for improving field work. The first column of Table 1 summarizes the general types of challenges discussed above (e.g., replication). The second column suggests some techniques that would speak directly to potential criticisms. The third

column offers some ways in which researchers can signal honesty and competence even if they are not able to demonstrate these qualities directly. Although such signals would not inherently alleviate the problem, they would provide audiences with greater assurance that work has been done well.

Table 1: Remedial measures

Challenges	Proposed remedial measure	Signals of competence
Training and professionalization	<ul style="list-style-type: none"> • Take courses on field work in doctoral programs, professional associations, and institutes. 	<ul style="list-style-type: none"> • Cite published work on field work techniques. • For graduate students, have advisors attest to advisees' skills.
Standards for interpreting evidence and adjudicating among competing propositions	<ul style="list-style-type: none"> • Identify interview questions that could yield evidence for alternative explanations. • Specify what sorts of responses or observations constitute confirmatory versus disconfirmatory evidence. • Provide an explicit method for coding responses from informants. 	<ul style="list-style-type: none"> • Discuss each informant's "leverage" when information from that individual is presented. • Provide assessments of each informant's candor.
Transparency in data collection methods	<ul style="list-style-type: none"> • Identify the sampling frame and selection strategy for informants. • Release interview guides or coding sheets (including prompts used). • List potentially relevant informants who were not interviewed and thus could be contacted as part of a replication exercise. 	<ul style="list-style-type: none"> • Discuss problems encountered in the field and remedial measures taken (e.g., in a lengthy Methodological Appendix).
Transparency about and potential replication of data analysis	<ul style="list-style-type: none"> • Release as much raw material from field work as confidentiality permits. 	<ul style="list-style-type: none"> • Whenever evidence from interviews is presented, report the number of informants asked each question and how they responded. • "Active citation" (Moravcsik 2014)
Double-blind controls in data collection and analysis	<ul style="list-style-type: none"> • Interview some subjects by proxy. • Have a third party review notes from the field to validate that the inferences drawn are correct. 	<ul style="list-style-type: none"> • Provide recordings of interviews (or of the researcher's side of the interview) to show that the researcher did not prompt informants to confirm the researcher's priors.

In terms of training (the first row in Table 1), doctoral programs might require that students take certain classes or modules on the craft of field work (rather than simply on research design). Researchers could also obtain some form of credential from professional associations, such as short courses at professional association meetings. In the absence of actual certification, advisors might informally certify (e.g., in recommendation letters) that graduate students possess certain skills. (This step would, of course, have the salutary effect of encouraging advisors to question students about how they conducted their field work and to help train them in field work methods.) Finally, researchers themselves can also signal that they are autodidacts in the method by discussing the relevant literature of field work in the context of a larger description of their method.

This larger description – normally, a Methodological Appendix – would present the researcher’s field work approach, discuss any problems she encountered, detail how she addressed those problems, and note any tradeoffs that she was forced to make between practicality and what might seem to be an idealized “scientific” approach. The notion of field method appendices is hardly novel: a number of prominent researchers in the discipline have carefully detailed the way they conducted their field work (e.g., Fenno 1978, Heard 1950). Inexplicably, however, fulsome discussion of what exactly was done is not yet a uniform practice in the discipline.

Field researchers can also take a number of steps to demonstrate that they have adhered to a clear and reasonable standard for adjudicating among rival interpretations of the evidence. For instance, they could identify which questions in their interview guides were designed to test alternative propositions or to deliberately provide interviewees the opportunity to offer evidence that might disconfirm their priors. They could then be explicit about what sort of responses to each of these questions constitute adequate evidence. In many cases, of course, it should be possible to quantify field work data and invoke ordinary standards for hypothesis testing. For instance, interviewers could compare the frequency with which respondents answered “yes” versus “no” to a particular question and report whether the difference was significantly different (in a statistical sense). The same is obviously true, of course, for coding of observations.

One novelty in Table 1 is the notion of discussing informants’ “leverage”. Some prospective interviewees have dispositive information about a particular incident (e.g., they were present at a meeting); others have less of a claim (e.g., they saw a transcript of the meeting, they talked to someone who was in a meeting immediately after that individual emerged from the meeting but were not there themselves, they talked to a participant in the meeting but only many weeks later, they heard about the meeting third-hand, etc.). Individuals with direct knowledge of events constitute high-leverage informants; what they say deserves more weight than those who can provide only hazy accounts or second-hand information. Likewise, some potential interviewees have stronger incentives to dissemble than others; the accounts of those who have no incentive to misrepresent what occurred deserve greater weight. By being explicit about their criteria for finding a particular report credible, researchers can signal they are holding themselves to a reasonable standard of empirical verification.

A compelling study should include a discussion of the sampling frame for interviews – ideally, the universe of potential interviewees and the strategy used to select among them (random, purposive, or some combination), as well as what replacements were identified for any selected informants who could not be contacted. This list could be supplemented by a list of potential informants not interviewed, so that a skeptic could decide for himself if there were some egregious omission, and so that another

scholar who wished to verify the findings could do so by talking to a sample of the uncontacted potential interviewees, without having to re-interview everyone.

Obviously, researchers should also make available their interview guides, coding sheets for observations, and other materials. If recordings have been made, the researcher can at least provide a recording of *her side* of a sample of the conversation, so that her skill in conducting the interview and her use of prompts can be evaluated (even if the other side of the interview cannot be made public).

As noted above, replication of data analysis requires that another scholar be able to pour over all material collected in search of evidence that might support a different conclusion. To that end, researchers should be prepared to make available as much of their data as they can without compromising informants' confidentiality. Video-recordings are obviously the "best evidence" for this purpose, as they capture all elements of the interview, followed in descending order of value by (b) audio recordings, (c) written transcripts, (d) redacted transcripts, and (e) summaries of the salient points of each interview. Where recordings are not possible, transcripts should include a description of the circumstances under which each interview was conducted, as these might inform interpretation of the informants' responses. For instance, a researcher reviewing someone else's field notes might discover that all informants who provided evidence for a particular proposition had been interviewed at their workplace during business hours, but notes from interviews conducted in settings where the subjects might have felt more relaxed provided evidence for different propositions. (One example along these lines comes from Mayhew [1974], whose initial interviews suggested that the motivations of U.S. Representatives were highly varied, but whose conversations with legislators many months later, after they had become accustomed to his presence, indicated that they cared almost exclusively about reelection.)

Field researchers can also be transparent in the way they present their "data". One possibility, proposed by Moravcsik (2014) would be to create a series of hyperlinks to more detailed documents for each bit of evidence presented. A less demanding approach would be for researchers to at least report all responses by interviewees to whom they put the same question (not just confirmatory or cherry-picked responses) in the text, footnotes, or an appendix. For instance, every time field researchers quoted a response by one of their informants as evidence, they could report the number of other informants to whom they put the same question and what fraction gave the same response.

In most field work, double-blind controls are missing at two levels: first, when interviews are conducted or observations are made, and second, when researchers write up that material. The most obvious way to deal with this challenge in the context of data collection is through proxy interviewing (Cammett 2013). But interviewers can often signal that they have not "led the witness" by being explicit about the prompts they used and providing recordings of their interviews (as discussed above). With regard to interpretation of the evidence, it is often possible to do a cursory version of "proxy write-up": that is, to have a third party review a sample of interviews and then write up conclusions based on them, which could be used to validate the original researcher's own findings. Such a review is, of course, immensely facilitated whenever at least some of the informants' responses can be coded systematically or text-analyzed.

Getting from here to there

The steps discussed above represent something of an ideal, which field researchers can only approximate most of the time. For instance, graduate students rarely have the resources to get others

to do their field work for them, and even if they could, proxy interviewing might have significant disadvantages that outweigh the benefits of double-blind controls. Likewise, confidentiality may impose radical limitations on what material can be made available for potential re-interpretation after the fact.

There is also likely to be a tension in practice between adherence to an interview guide (in order to ensure that different informants are asked the same questions in the same way) and the opportunity to explore unexpected and intriguing pieces of information that emerges in the course of an interview. This tradeoff is less acute, of course, when interviews are being conducted purely for the purpose of formulation propositions or purely for the purpose of testing them, rather than a combination of both at the same time.

Although practical constraints will inevitably create some tradeoffs and tensions, many of the measures discussed above are not costly to adopt. For instance, it should be easy for researchers to begin their field work preparations by (a) specifying the ideal universe of informants whom they would wish to interview, (b) articulating their selection procedure, and (c) specifying replacements for those interviewees in the ideal set who cannot be contacted. (Indeed, this process may be conceptually clarifying for researchers even if they cannot rely exclusively on the resulting sample for practical reasons.) Likewise, it is not excessive to insist that researchers to make available interview guides, articulate their standard for confirming a hypothesis, make available at least some notes, write up their findings in a transparent way, and provide a detailed methodological appendix.

In terms of training, researchers can certainly take courses in field work method – even if taught in other disciplines – that give them some practical exposure to the techniques they will use. In addition, at least two volumes have recently been published that provide valuable insights into the how to conduct field work in political science (Mosley 2013, Kapiszewski et al. 2015; see also Bennett and Checkel 2014 and Gusterson 2008); these supplement classic works on ethnographic research from both political science and anthropology (*inter alia*, Emerson et al. 2011, Wood 2007, Van Maanen 2011). Finally, some advisors may be willing to accompany graduate students on an interview or two, or at least to practice an interview with them. In short, there is really no reason for researchers to go into the field ill-trained.

For the more “expensive” elements of Table 1, one strategy researchers might consider would be to request funding for these proposed measures in their grant proposals. Although funding may prove difficult to obtain, including such a request in a proposal indicates seriousness of purpose and methodological care to reviewers of the proposal. In addition, those denied funding have at least sent a signal that they take methodological issues seriously.

A less expensive option is for peers to enlist each other in validating their findings. For instance, a researcher might pass notes that could not be made fully public to a graduate student, a TA, or an advisor for verification that the material had been interpreted reasonably. Graduate students might also apply the same strategy for proxy interviewing: e.g., an advisor or fellow student might interview one or two subjects who were not in the researcher’s sample, to see if they confirmed what other sources had reported. Such steps are not totally “free” in terms of time, but they are trivial in terms of the total time dedicated to a project, and they do not require additional funding.

An example

To illustrate how researchers might incorporate some of these elements into their work, it may be useful to consider a specific example. Here I draw on a study of science policymaking in the United States conducted by a graduate student under my direction.

Table 2: Before and after

	Original proposal	Revised work plan
Training	<ul style="list-style-type: none"> • Informal conversation with advisor. • Read one article on interviewing. 	<ul style="list-style-type: none"> • The student took a research methods class, did additional reading, and strategized how to conduct the interviews with his advisor. • The advisor and the student jointly interviewed one individual not in the original sample (see below) to practice technique and to verify the findings from other interviews.
Interviewee selection	<ul style="list-style-type: none"> • Notional list of potential interviewees, with a potential snowball sample. 	<ul style="list-style-type: none"> • The researcher identified a universe of ~160 potential interviewees, consisting of all individuals who had held specific positions in the relevant agency over the previous 20 years. From these, a random sample of 15 junior staffers (total N = ~125) and a purposive sample of 15 high-leverage senior staffers and political appointees (total N = ~35), were selected. Reasonable replacements were identified for senior officials who could not be reached (e.g., a principal's chief of staff instead of the principal).
Method	<ul style="list-style-type: none"> • Rough interview guide. 	<ul style="list-style-type: none"> • An interview guide was developed and pre-tested on individuals not included in the sample, and made available for review. • Items in the interview guide were classified ahead of time according to the proposition they were designed to test. • A field work method section was added as an Appendix.
Interviewing	<ul style="list-style-type: none"> • No special steps taken. 	<ul style="list-style-type: none"> • A brief description of the context of each interview was included in summary notes. • One interview was recorded, with the interviewee's side of the conversation deleted, to demonstrate the researcher's craft. • The advisor and the student jointly re-interviewed two high-leverage informants.
Data	<ul style="list-style-type: none"> • No specific plan. 	<ul style="list-style-type: none"> • All interview notes were redacted to protect informants' confidentiality, with extensive material available for review.
Write-up	<ul style="list-style-type: none"> • "Tell me a story" 	<ul style="list-style-type: none"> • Each assertion based on an interview was accompanied by a note reporting the number of individuals who were asked the same question and how many gave similar answers. • Notes were briefly reviewed by the advisor as a validation exercise.

The original proposal for this project approximated the “soak and poke” approach which characterized earlier generations of field work (Fenno 1986), and which would have been fine if the goal of the field work was to simply develop priors. In this case, however, the goal was to collect evidence for or against fairly well-defined propositions about the role a particular agency played in the budgeting process. Although the revised version of the research design did not include every suggestion in Table 1, it was more systematic. Table 2 presents the “before/after” comparison.

The “after” picture is, of course, imperfect. In terms of training, it would have been better if the student in question had taken a short course at the Institute for Qualitative and Multi-Method Research or some other formal training. Likewise, because all informants had been ensured confidentiality, and the information that interviewees provided would inevitably identify them, some of the redactions were draconian (e.g., four pages of interview notes were reduced to a few sentences). Third, the validation exercise was cursory. Finally, some of the more ambitious elements had to be scaled back in practice (for instance, some selected interviewees for whom there was no obvious replacement declined to be interviewed). Nevertheless, the revised research design provided a much better justification for the “professional presumption” than is often the case.

The revised method also offered corollary benefits. Interviewees – some of whom had gone to graduate school themselves – were impressed by the approach. For the advisor, the joint interviews conducted with the student were gratifying and enjoyable without being particularly time-consuming. In a context of increased scrutiny of graduate students’ claims, it was also reassuring when re-interviews echoed the student’s original findings.

Most important of all, the process of revising the field work plan was extremely clarifying for the student. The rationale for each interview and the sort of information that needed to be covered became much more apparent. Particularly revelatory was the process of thinking through what sorts of responses from interviewees would be required to support which hypotheses: an earlier draft of the original interview guide, which had looked quite thorough, was found to be missing questions that would permit the student to rule out certain alternative explanations.

Conclusion

Field researchers, including graduate students, frequently conduct dozens of interviews over the course of many months. Unfortunately, because of the problems discussed above, this research often boils down to a few quotations on a slide in a job talk, a section of a published article, or scene-setter chapters in a book manuscript. The evidence they do present may or may not be taken seriously. Given the amount of energy researchers invest in field work, they should get a higher return.

This essay suggests some ameliorative approaches that collectively cover almost all elements of field work: researcher training, selection of informants, interviewing, write-up, etc. There will inevitably be tradeoffs between the desire to maintain “scientific” standards and the practicalities of field work, and in the end, researchers will have to weigh each choice. But even that process – imagining the methodologically optimal way to conduct field work and then justifying the concessions made on practical grounds – would elevate the standard of research.

References

- Bennett, Andrew and Jeffrey T. Checkel, eds. 2014. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge University Press.
- Cammett, Melani. 2013. "Proxy Interviewing as a Research Strategy." In Layna Mosley, ed., *Interview Research in Political Science*. Cornell University Press.
- Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 2011. *Writing Ethnographic Fieldnotes* (2nd edition). University of Chicago Press.
- Fenno, Richard. 1978. *Home Style: House Members in their Districts*. Little, Brown.
- Fenno, Richard. 1986. "Observation, Context, and Sequence in the Study of Politics", *American Political Science Review*, 80 (1): 3-15.
- Gusterson, Hugh. 2008. "Ethnographic Research." In Audie Klotz and Deepa Prakash, eds., *Qualitative Methods in International Relations: A Pluralist Guide*. Palgrave Macmillan.
- Heard, Alexander. 1950. "Interviewing Southern Politicians", *American Political Science Review*, 44: 886-896.
- Loeza, Soledad. 2005. "Off to the Field." *APSA-Comparative Politics Newsletter* 16(2): 9-12.
- Kapiszewski, Diana, Lauren M. MacLean, and Benjamin L. Read. 2015. *Field Research in Political Science: Practices and Principles*. Cambridge University Press.
- Mayhew, David R. 1974. *Congress: The Electoral Connection*. Yale University Press.
- Moehler, Devra C. 2005. "Off to the Field." *APSA-Comparative Politics Newsletter* 16(2): 16-8.
- Moravcsik, Andrew. "Transparency: The Revolution in Qualitative Research", *PS*, January: 48-53.
- Mosley, Layna ed. 2013. *Interview Research in Political Science*. Cornell University Press.
- Stevenson, Randolph T. 2005. "Making a Contribution: The Role of Fieldwork in Scientific Research Programs" *APSA-Comparative Politics Newsletter* 16(2): 13-15.
- Van Maanen, John. 2011. *Tales of the Field* (2nd edition). Chicago University Press.
- Wood, Elisabeth. 2007. "Field Methods." In Charles Boix and Susan Stokes, eds., *Oxford Handbook of Comparative Politics*. Oxford University Press.

MIT OpenCourseWare
<https://ocw.mit.edu>

17.801 Political Science Scope and Methods
Fall 2017

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.