ARTICLE

Ethnography

http://eth.sagepub.com Vol 10(1): 5–38[DOI: 10.1177/1466138108099586]

'How many cases do I need?'

On science and the logic of case selection in field-based research

■ Mario Luis Small

University of Chicago, USA

ABSTRACT ■ Today, ethnographers and qualitative researchers in fields such as urban poverty, immigration, and social inequality face an environment in which their work will be read, cited, and assessed by demographers, quantitative sociologists, and even economists. They also face a demand for case studies of poor, minority, or immigrant groups and neighborhoods that not only generate theory but also somehow speak to *empirical* conditions in *other* cases (not observed). Many have responded by incorporating elements of quantitative methods into their designs, such as selecting respondents 'at random' for small, in-depth interview projects or identifying 'representative' neighborhoods for ethnographic case studies, aiming to increase generalizability. This article assesses these strategies and argues that they fall short of their objectives. Recognizing the importance of the predicament underlying the strategies – to determine how case studies can speak empirically to other cases – it presents two alternatives to current practices, and calls for greater clarity in the logic of design when producing ethnographic research in a multi-method intellectual environment.

KEY WORDS ■ ethnographic methods, generalizability, representativeness, validity, case study, sequential interviewing, extended case method, science

Probably the most memorable of the lectures of Nobel-prize winning physicist Richard Feynman was his 1974 commencement address at Caltech, where he described what he calls 'cargo cult science'. Feynman, worried about the preponderance of what he believes are pseudo-sciences, compares these practices to the cargo cults of the South Pacific:

> In the South Seas there is a Cargo Cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in ... [the controller] – and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things Cargo Cult Science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land. (Feynman, 1999: 208–9)

Feynman's lecture was devoted largely to practices such as ESP, but the hints about social science were difficult to miss. (In fact, he criticized a psychologist's advice to his student not to replicate the studies of others.) Pseudo-scientists were expert imitators but terrible practitioners, adopting the form of science but nothing else. In this respect, his analogy might have been stronger had he noted that some New Guinean cargo cultists had fashioned their own airplanes out of logs, sticks, and leaves, remarkably accurate replicas that, lacking engines and a foundation in aerodynamics, would never fly (Harris, 1974; Worsley, 1968).

While Feynman probably underestimated the successes of social science, his observations are worth noting by at least one major segment of contemporary ethnographers, for whom the temptations of imitation have never been stronger. The problem of imitation is not new to social scientists, who from the start have argued heatedly (and repeatedly) over the merits of emulating the natural sciences in pursuit of social scientific methods (Dilthey, 1988; Lieberson and Lynn, 2002; Saiedi, 1993). But today, an important subset of ethnographic researchers – and of qualitative researchers more generally – faces its *own* version of that dilemma: whether to emulate basic principles in quantitative *social* sciences in establishing standards of evidence for qualitative work. Some background is necessary.[1]

The predicament of ethnographic work in multi-method contexts

The predicament arises from what might seem to be an unqualified accomplishment, the simmering of the counterproductive debates, which reached a boiling point during the 1980s, over the relative merits of quantitative versus qualitative research. Today's calmer waters have been especially regenerative for the fields of urban poverty, social inequality, and immigration, where both quantitative and qualitative works flourish, and where experts in one methodological tradition frequently cite those in others. In fact, several major studies in these fields have employed, with varying degrees of integration, both quantitative and qualitative data, the latter being at times interview-based and at times ethnographic. Examples are Portes and Rumbaut's recent studies of the children of US immigrants (Portes and Rumbaut, 2001; Rumbaut and Portes, 2001), Wilson and his colleagues' studies of urban conditions in Chicago (Wilson, 1996; Wilson et al., 1987), and England and Edin and their colleagues' studies of urban single mothers in the US (see England and Edin, 2007). These fields stand in contrast to others where, for epistemological, political, or historical reasons, most practitioners work within a single method or set of methods, as in symbolic interactionism or the interpretive work on culture.

But the more cooperative spirit in urban poverty, social inequality, and immigration has only spread so far. Despite the more methodologically open environment, research in these fields remains dominated by quantitative sociologists, demographers, and even economists (Wacquant, 1997; see also Burawoy, 2005). While important ethnographies in these fields continue to be published and highly cited (e.g. Duneier, 1999; Levitt, 2001; Pattillo, 1999; see Newman and Massengill, 2006), most articles in these fields published in the top generalist journals, such as *American Journal of Sociology*, the *American Sociological Review*, and *Social Forces*, remain quantitative in nature.[2] The preponderance of statistical research stems in part from the steady and continuous supply of easily accessible quantitative data, such as the decennial US Census, the Current Population Survey, the Panel Study of Income Dynamics, and the National Longitudinal Survey of Youth, all of which contain many variables related to poverty, immigration, urban conditions, neighborhoods, and socio-economic status. In addition, urban poverty, social inequality, and immigration (along with criminology, education, and public health) remain among the most highly funded fields in US social science, and the largest funders in these fields – nonprofit organizations and government agencies such as the Russell Sage Foundation, the Ford Foundation, the National Science Foundation, the US Department of Housing and Urban Development, and the National Institutes of Health – have exhibited a greater inclination to fund quantitative projects. The edited volumes generated by these funding streams bear evidence to this claim. Consider a few highly cited collections: Goering and Feins' (2003) recent volume on the effects of neighborhood poverty, funded by HUD; Neckerman's (2004) volume on social inequality, funded by Russell Sage; and O'Connor et al.'s (2001) study of urban inequality in multiple cities, funded by both Russell Sage and the Ford Foundation. Each of these volumes, while demonstrating deference to ethnographic research, overwhelmingly reports quantitative findings.[3]

These circumstances have produced at least two consequences. First, in contrast to other fields where ethnographers work, ethnographic studies in urban poverty, inequality, and immigration are often evaluated – even on methodological grounds – by quantitative researchers. Thus, while ethnographers doing work on, for example, narrative and culture can confidently expect their work to be reviewed primarily, if not exclusively, by other qualitative researchers, those in the aforementioned fields must count among the potential reviewers of their work, demographers, quantitative sociologists, and even economists – scholars who are experts on the *subject matter* without necessarily being experts on the *method*. Inevitably, the reviews will cover issues not traditionally addressed in closed ethnographic intellectual communities. For example, my *Villa Victoria*, an ethnographic study of a predominantly Puerto Rican housing complex in Boston, was reviewed in *Contemporary Sociology* by a demographer, who, in a generally balanced article, covered both theoretical and methodological issues. Yet the methodological discussion did not center on the extent to which I attained an empathetic understanding of my informants, on the level of reflexivity in the work, or on the extent to which the history of the neighborhood informed the analysis; instead, it focused on one of demography's central concerns, whether the neighborhood was representative:

> While the conditional approach is an asset from a theoretical stand point, it is a limitation from a methodological perspective. The generalizability of the findings is a concern. Since theories are based on generalizations, focusing on exceptions as advocated by the conditional approach makes us question the applicability of these findings to other neighborhoods. (Morales, 2006: 284)

Consider another example. Fordham and Ogbu's (1986) highly cited study of 'acting white' among black students in a Washington, DC middle school has been rediscovered over the last 10 years by sociologists and economists interested in culture and urban inequality, a rediscovery oriented toward methodological and empirical issues. But this attention centered on none of the core issues identified in Burawoy's (2003) recent review of types of ethnographic 'revisits', such as reconstructing the theory after returning to the school or assessing the structural forces at play in US cities during the late 1980s. Instead, the new critiques focused on whether the few dozen students interviewed by Fordham and Ogbu were statistically representative of black students in US society (Ainsworth-Darnell and Downey, 1998; Cook and Ludwig, 1997). Consider a third example, from the literature on transnationalism, where ethnographies, such as Levitt's (2001) study among Domincan migrants to the US, have confirmed the prevalence of transnational practices among immigrants in all continents. In an important recent study evaluating these works, quantitative researchers critiqued the latter on methodological terms – but not by focusing on the authors' ability to conduct multi-site ethnography, or on the challenges of establishing rapport quickly and effectively in different countries. Instead, the critiques focused on the questions that worry demographers, even employing explicitly the language of variable selection:

> This emerging literature on immigrant transnationalism is characterized by an empirical base consisting exclusively of case studies. Qualitative case studies *consistently sample on the dependent variable*, that is, they document in detail the characteristics of immigrants involved in transnational activities but say little about those who are not. (Portes et al., 2002: 279; italics added)

We may set aside for the moment whether any of these critiques is valid in order to highlight the larger predicament: ethnographers in these fields can expect to be assessed on their methods by quantitative researchers.

Second, ethnographic studies in urban poverty, social inequality, and immigration operate in an environment at once thirsty for in-depth case studies that describe conditions in poor, minority, immigrant, or otherwise non-mainstream groups, neighborhoods and communities, *and* either skeptical or uncertain about the relationship between these 'small-*n*' studies and the larger population of groups, neighborhoods, and communities that the case studies are expected to represent (Lieberson, 1991). Consider a contrast. When Geertz (1973) wrote on the cockfights in a small Balinese village, many expected his *theoretical model* (of how games can embody societal power relations) to be applicable to other sites, but few expected the *empirical findings* to be so applicable – that is, for cockfights to look similar or to follow the same rules in other villages throughout or outside of Indonesia. The latter would be wholly beside the point.[4] But ethnographers in the aforementioned fields work in a professional, intellectual, and policy environment that demands empirical findings applicable to other cases. When ethnographers today describe conditions in, for example, one poor black neighborhood in St Louis, MO, many of their readers in the urban poverty literature expect to be learning about the conditions in poor black neighborhoods in general – in Boston, Los Angeles, New York, and perhaps even London and Rio de Janeiro – not merely that neighborhood in St Louis. In fact, ethnographies of single neighborhoods are often cited to argue that conditions in 'the ghetto', 'the underclass', 'the immigrant enclave', or other similar categories exhibit some set of *empirical* characteristics, such as desolation, graffiti, street disorder, gang presence, street entrepreneurship, etc. (on the perils of this practice, see Small, 2007, forthcoming b). Similar expectations surround the interpretation of in-depth interview studies with small numbers of respondents. When interpreting a study of, for example, 40 immigrant low-income women in San Diego, CA, researchers expect to learn something *empirical* about the conditions of low-income immigrants in other cities and regions, not merely about those 40 women. In sum, in fields such as urban poverty and immigration in the US, ethnographic case studies are often explicitly expected to represent larger entities. Nevertheless – and this is key – the precise relationship between such cases and the larger populations they are expected to represent remains unclear.

It is in this context that imitation, as described earlier, arises. The core predicament that many ethnographers face is deceptively straightforward: how to produce ethnographic work that keeps at bay the critiques expected from quantitative researchers while also addressing the thirst for in-depth studies that somehow or other 'speak' to *empirical* conditions in *other* cases (not observed). How to select studies such that these types of inferences are valid?

As I discuss below, for all too many in recent years, the answer has been to emulate the practices of quantitative researchers, especially the foundation in classical or frequentist statistics – that is, to match, as closely as either possible or practical, standard survey practices designed to ensure representativeness, and thus generalizability (e.g. Klinenberg, 2002). The result has been a form of imitation grounded in language, or, more precisely, the adoption of words with only a superficial (and at times incorrect) application of their meaning. Today, many researchers in these fields encourage their students doing ethnographic work to ensure that their small samples are 'unbiased' or that their single-case studies are 'representative' or 'not selected on the dependent variable'. As we shall see, the most formal expression of this perspective is represented in King et al. (1994), but informal versions of this approach abound. The basic critique of this article is that these practices constitute little more than applying words without adopting their meaning, constructing sticks-and-leaves airplanes that will never fly.

The problems faced by ethnographers working in interdisciplinary fields today are both substantively and publicly important, creating a need for programmatic clarity about these questions. That clarity, however, has been missing, and, in fact, many of the most commonly employed fieldwork manuals hardly address the topic (Hammersley and Atkinson, 1995; Lofland and Lofland, 1995).[5] Methods of scientific inquiry are languages to the extent that they constitute systems of thought, with terms and ways of framing problems that are specific to their systems. I will argue that solutions should involve developing alternative languages and clarifying their separate objectives, rather than imitating the language of classical statistics for problems to which it is not suited. What I have described is a large and complex problem that I cannot hope to cover entirely in one article. Instead, I will examine one issue, that of 'generalizability', one of the most central concerns that quantitative researchers raise about ethnographic research in the aforementioned fields. To make the problem clear, I will discuss two common scenarios, one from in-depth interview work and one from participant observation, under which ethnographers concerned about science attempt to make their work 'generalizable'. After critiquing these forms of thinking about generalizability, I discuss two major alternatives.[6]

First scenario: the 'representative' small-*n* interview-based study

One of the most common practices in qualitative sociology today is to conduct in-depth, open-ended interviews. Sometimes, these are fielded on hundreds of respondents (Lamont, 1992; Newman, 1999); more typically, they involve a small number of interviewees (30 or 40). It is in this latter case that the problems are more salient, since the questions of generalizability are more obvious and the answers to these less clear. Consider the following scenario.

The most common question

Jane, for her second-year paper, would like to study the attitudes toward immigration of working-class African Americans. To capture the depth and nuance of these attitudes, she wishes to conduct lengthy open-ended interviews, and she is planning to interview 35 respondents. Having taken courses in basic statistics and research design, Jane worries that, by doing 'qualitative work', she is not conducting truly scientific research. She sits before her adviser, and asks one of the most common questions posed by beginning students doing in-depth interview work in these fields: 'How do I ensure my findings will be generalizable?'

Her adviser 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 percent 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. (Experienced interviewers know these figures to be highly optimistic.) The city is so segregated that all of them happen to be African American. She conducts 35 high-quality two-hour interviews that delve deeply into attitudes about immigration, uncovering subtle relationships among attitudes, experience with discrimination, gender, and Southern origins, and she happily completes her paper. Since (she believes) her method mirrors that of Lamont's (1992) well-regarded *Money, Morals, & Manners*, she is confident in the 'generalizability' of her findings about the black working class.[7]

Jane's predicament is standard; her solution, rather common (e.g. Hermanowicz, 1998; Walzer, 1997; to a lesser extent, Aries and Seider, 2005). There is, however, an important problem: under no statistical definition of 'generalizability' could the responses of those 35 individuals be considered to reflect reliably the conditions of the African American working class. From the perspective of statistical generalizability, Jane's research suffers from two problems.

First, the sample has an inbuilt and unaccounted for *bias*.[8] Jane only interviewed the 35 percent of respondents who were polite enough to talk to her, friendly enough to make an appointment based on a stranger's cold call, and extroverted enough to share their feelings with this stranger for two hours. These people may have systematically different attitudes about others, including immigrants, than non-respondents. Since she knows nothing about those who did not respond (they hung up), she has no way of adjusting the inferences she obtained from the 35 respondents. In addition, since she knows nothing about working-class blacks in *other cities*, she still does not know if, had she attained 100 percent response rates, the respondents in her city would be 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. Among my graduate students working on in-depth interview study designs, perhaps the single most common question I hear is, '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 factors (King et al. (1994) 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 discover how many working-class blacks are pro-immigration reform (one 'yes or no' question); and that she wanted to be confident at a 95 percent level that the average in her sample matched the average in the population at large within +/– five percentage points; and that the population of working-class blacks in the US was 2,000,000 people (for large populations, the exact size does not matter very much).[9] In this case, she would need 385 respondents.[10] If Jane narrowed her focus, and only wanted to be confident about the 1000 working-class blacks in one city, she would need 278.

In fact, nothing Jane could do with her three dozen people would ensure that they matched, with high certainty, the characteristics of the US working-class black population. Some interview-based researchers prefer to ignore these two problems and refer to studies such as Jane's as 'representative' because they were based on a 'random' sample. In doing so, they are simply adopting words without adopting their meaning, referring to wooden tubes as airplanes. If Jane sought a representative study, she should have conducted a survey (and assumed all of its limitations, such as having to construct a formal questionnaire and reducing its length to maintain high response rates).

Better than nothing?

The natural question is whether, having acknowledged these difficulties, Jane is still not better served by selecting her respondents 'at random' (in quotation marks because her final sample is not statistically random due to the large proportion of people who did not respond) than to have adopted some other non-random selection strategy. Not always. Consider two standard alternatives in light of these questions.

One is *sampling for range*, a longstanding technique which might yield better and more reliable data depending on her interests (Weiss, 1994). By this technique, the analyst identifies sub-categories of the group under study and ensures to interview a given number of people in that category. For example, suppose Jane suspected that gay and lesbian respondents might be more sympathetic to immigrants, and decided this was one of the core issues she wanted to understand. Even a truly random sample would have yielded, at best, three or four gay or lesbian respondents out of 35, of which one or two, at best, would reveal this to her. This would leave her no room to examine her 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 required finding them through non-random means, such as contacting formal organizations and requesting references. For many questions of interest to interview-based researchers in the social sciences, sampling for range is more effective. Of course, from the language of statistics, Jane will continue having problems with bias and representativeness. However, from a more appropriate language, she will now be able to identify (at least some of) the mechanisms through which sexual orientation and attitudes about immigration are related.

Another alternative is *snow-ball sampling*, the well-known practice of asking interviewees to recommend other interviewees (Weiss, 1994). Snowballing almost always increases the number of respondents, because people become more receptive to a researcher when the latter has been vouched for by a friend as trustworthy. (This trustworthiness might also translate into greater openness, producing deeper interviews, but I have not seen anyone demonstrate this proposition.) For example, if Jane had contacted 100 people, each based on a personal referral, she might have interviewed 75 (it is impossible to know), but certainly many more than the number obtained by cold-calling.

One consequence of snowballing is that the final interviewees are more likely to know one another than would be the case had they been selected at random. Thus, they are more likely to constitute a social network. For this reason, many argue, again following the precepts of classical statistics, that snowball samples in a small study such as Jane's would be 'more biased' than the 'random' sample she employed. (This type of critique is often applied to interview-based studies of social networks.)

I believe this view is mistaken on two grounds. First, suppose we assessed both methods by the standards of classical statistics. In one case, there are 75 out of 100 people interviewed through personal referrals; in the other, there are 35 who agreed to participate out of 100 selected at random. By the standards of classical statistics, *both* samples would have a bias problem, the first due to in-network selection and the second due to non-response. What proponents of the random selection approach to small-*n* in-depth interviewing rarely mention is that many people who are cold-called will not agree to long, in-depth interviews on personal topics with a stranger. This often buried detail – how many people refused, hung up, or were not home? – is critical. I am not surprised that many such studies do not report it.

Second, we should not assess *either* method by the standards of classical statistics. That is to say, in a deeper and more important sense, 'bias' is the wrong term. What an in-depth interviewer with three dozen respondents faces is not a 'bias' problem but a set of cases with particular characteristics that, rather than being 'controlled away', should be understood, developed, and incorporated into her understanding of the cases at hand (more on this below). By inaccurately labeling 'representative' the strategy that Jane employed, we erroneously assume that the others are 'biased'. Not only would Jane's ostensibly representative sample be biased by quantitative standards – neither bias nor representativeness is an appropriate standard for the kind of research Jane wanted to conduct.

To be clear, there are only two solutions to Jane's predicament: she should not conduct the study or she should change the language through which she interprets it. The half-way solution, of 'trying to make it representative', does not work. Many are reluctant to agree to the first because it seems to imply that all interview-based studies should involve large samples (such as Newman, 1999). While such large studies are valuable, there is a place for a small interview study to make meaningful contributions to knowledge, provided the language and assumptions through which it is interpreted differ. In fact, as I suggest later, we might benefit from not calling Jane's interviewees a 'sample of n = 35', but, instead, a 'set of 35 cases'.

Second scenario: the study of an 'average' neighborhood

There is a similar problem in participant observation research aimed at dealing with large-n questions. Today, major generalizations in both the public and in the academic sphere are made about conditions in urban areas, such as black and Latino inner-city neighborhoods in the US, *favelas* in Brazil, or *banlieues* in Paris. Ethnographers are expected and asked to penetrate the neighborhoods and report to the public on conditions in such neighborhoods, not only to test the generalizations but also to produce their own, to explain, for example, whether neighborhood conditions exacerbate the effects of heat waves or natural disasters. The question remains – how to select a case and design a study to render such statements empirically valid? Again, a common answer, implicit or explicit, is to search for representative cases. A hypothetical scenario should be instructive.

Selecting the average

Bill, a graduate student in sociology, wishes to study how neighborhood poverty affects out-of-wedlock births, by conducting an in-depth ethnographic study of a single high-poverty neighborhood. His main concern is to uncover the set of mechanisms underlying this process. However, like Jane, he wants to ensure his findings are 'generalizable' to poor neighborhoods. Dissatisfied with the absence of a clear answer in his standard graduate methods course, he turns to the highly citied King et al. (1994) model. The authors address a similar problem directly, and Bill finds solace in their solution:

> 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 ... (King et al., 1994: 67–8)[11]

Bill turns to the census, finds a neighborhood that is 40 percent poor, 60 percent black, with 80 percent 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 (or, to paraphrase King et al., an average poor community). To comply with IRB regulations, Bill gives his neighborhood a pseudonym, 'The Streets', and reports that it is located in a 'large Midwestern city'.[12] Bill conducts his study, and finds that residents in this neighborhood are distrustful of each other. This distrust, he finds, makes women unwilling to marry the fathers of their children. Since his neighborhood was representative, Bill surmises, he is confident that neighborhood poverty increases out-of-wedlock births in poor neighborhoods at large through the mechanism of reduced trust, among mothers, toward potential fathers.

Either explicit or implicit versions of Bill's approach to selecting study sites – that is, studying neighborhoods that seem to be typical ghettos, enclaves, or housing projects in order to improve one's confidence in the generalizability of empirical statements to other neighborhoods – abound.[13] The problem with this approach is that, no matter how Bill selected his single neighborhood, it will never be representative of poor neighborhoods. Bill suffers from two problems. The first is exacerbated by his conception of an 'average' neighborhood. Since Bill lacked the means to determine – beyond the limited neighborhood characteristics recorded in the US Census – that his neighborhood was in fact statistically average, he settled on a neighborhood that *appeared* typical, one exhibiting a preponderance of graffiti, for example, since this accorded with his picture of the conditions in an 'average' poor neighborhood. This selection was no doubt influenced by film and television images of poor neighborhoods, images produced for commercial reasons which may or may not bear any resemblance to statistically average neighborhoods. The second problem is Bill's confusion of 'representative' with 'average'. The neighborhood's conditions may happen to match the traits that, from the census, one knows to be at the mean on observed demographic characteristics. 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 meet this criterion*. In fact, there is no way to determine whether the same results would have been obtained by studying the population of neighborhoods at large. Bill has painted himself into the figurative corner in which he has assumed a warrant to make generalizations based on a one-*n* study (Lieberson, 1991; on warrants, see Katz, 1997).

To be clear on the hopelessness of Bill's situation given his objectives, consider one potential solution, obtaining copious and very detailed information on the neighborhood, the generally sensible recommendation by King et al. (1994). Some believe that Bill could make up for the 'lack of variation' in his single study by generating variation from within, and make up for the overall 'lack of statistical power' by digging deeper into his single case. Bill, for example, could examine in depth his neighborhood's history and institutions, and produce a rich, detailed, multi-layered study of its conditions. Obtaining more data on a case is always a laudable practice, but from the perspective of statistical generalizability, this 'solution' offers no advantage. Suppose that instead of neighborhoods we were studying individuals, and selected one person with the characteristics of the average American: a married 36-year-old female with a high school education who earned $36,911 last year.[14] We interviewed this female for many 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 fact.[15]

The same logic is applicable in Bill's case. As his readers, we may want to believe that he tapped into the true (essential?) nature of 'the American ghetto' because the neighborhood either matched census-collected traits of the average poor neighborhood or it generally accorded with pre-existing notions of conditions in such neighborhoods, a sentiment reinforced by the neighborhood's pseudonym ('The Streets') and anonymous location. In fact, *this is precisely how researchers in urban sociology, immigration, and social inequality frequently interpret and cite neighborhood ethnographies* (see Small, 2007, 2008b). But reason would compel us to think otherwise: Bill has earned a warrant to speak empirically about the average 'ghetto' that is *no greater* than a study of the single 36-year-old female would have earned to speak empirically about the average American. In the language of statistics, these are both samples of one.

Better than the alternatives?

One could question, again, whether Bill's strategy is not superior to any alternative. An alternative is to search for *unique cases* (Small, 2004; van Velsen, 1978 [1967]). Suppose that Bill had chosen a neighborhood with a 40 percent poverty rate but little 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 assume that inferences based on this neighborhood would be less statistically generalizable than those based on 'The Streets', since the latter was more 'typical' (in fact, this assumption informed Morales, 2006). But, based on a sample of one, they are neither more nor less so. As before, one could ask if there is any harm in pursuing the statistics-inspired 'random' or 'average' strategy. Sometimes there is. Suppose the mayor in the second case also had instituted a radical and unique policy whereby mothers received significantly higher rent subsidies plus $1000 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. In case studies, rare situations are often precisely what the researcher wants (Small, 2004; Yin, 2002; also Schudson, 1992). (More on this below.)

Alternatives

Jane and Bill's designs, and related approaches to render field studies more generalizable, simply fail in their objectives. I see only three solutions to this predicament: 1) to ignore the problem; 2) to qualify the work by explaining that it is provisional or 'hypothesis generating'; and 3) to conceive and design the work from a different perspective and language of inquiry.

While ignoring the problem may seem to be inappropriate, it remains a suitable option under some circumstances, given the different goals that underlie different theoretical perspectives. As in the case of Geertz's (1973) cockfight study, the theoretical perspective informing a case study may render irrelevant whether the study provides empirical information on conditions in other cases. For example, ethnomethodologists aim to study practical life as it is experienced, and they assume that social reality is both inter-subjective and radically situational, such that the meaning of actions and events can only be explained within the particular set of circumstances in which they occurred (Garfinkel, 1967). This assumption (that meanings are indexed to particular situations) implies that whether other sites – for example, other neighborhoods – would exhibit the same empirical characteristics is not merely beside the point, but it would erroneously presume that the analyst can simply extract the characteristics of a neighborhood from the context in which actors experienced them and expect them to have the same consequences elsewhere. Similarly, narrative theories fundamentally assume that identity derives from personal experiences and the narratives actors tell themselves about them. And while conditions shared with others, such as the experience of race discrimination, can affect an actor's narrative of self, her ultimate narrative identity depends on her accumulation of experiences, which is hers alone (Somers, 1994). In this light, asking whether the experiences of one actor reveal something *empirical* about others would be conceptually senseless, and beside the point.[16]

But while these and other perspectives can afford to (in fact, by design often must) ignore the problem without consequences, not all perspectives can. The intellectual, professional, and public environment described at the start of this article undermines efforts to simply ignore the problem in the context of urban poverty, immigration, and social inequality scholarship. The field requires ethnographic and in-depth interview-based case studies, and it requires some answer about the empirical relationship between such case studies and other sites not observed.

The second solution, conceiving of the work as hypothesis generating, is also suitable. However, as I show below, hypotheses may be either logical or senseless, and different strategies will yield substantially better hypotheses than those a project such as Bill's might generate. Simply adding a qualifier or disclaimer without articulating the difference between good and bad hypotheses is merely a reactive solution to an external threat, not an epistemological position grounded in a sound perspective.

In the rest of the article, I make a case for the third option that provides a way of not only generating better hypotheses but also producing compelling empirical statements. While several of the ideas I propose have been discussed in other contexts, they have not been examined in light of the contemporary predicament presented at the start of this article. For this reason, the ideas require refinement, development, and in one case major elaboration; their implications for what scholars such as Jane and Bill should do must be articulated. I will discuss systematically two alternatives, an extension of the extended case method and what I term sequential interviewing.

First alternative: extending the extended case method

Probably the most prominent recent solution to the problem of generalization in case studies has been the extended case method, by which researchers analyze a particular social situation in relation to the broader social forces shaping it. In American sociology, the extended case method is most often associated with Burawoy (Burawoy, 1998; Burawoy et al., 1991); elsewhere, it is associated with Gluckman and the Manchester School (Gluckman, 1961; also Evens and Handelman, 2006). Even though Burawoy learned the method from van Velsen, one of Gluckman's students (Burawoy et al., 1991), there are still, in fact, several extended case methods, or several conceptions of the core of the procedure. For example, Mitchell (1983) and van Velsen (1978 [1967]) describe the core of the method in somewhat different terms, with the latter placing greater emphasis on the identification of unique cases; and Burawoy's latest (1998) formulation differs in important details from his initial statement in 1991, which itself was an 'extension' of van Velsen's (Burawoy et al., 1991: 326). For our purposes, two of these formulations matter most, Burawoy's and Mitchell's.

The Berkeley school

Michael Burawoy made his first major statement on the method in a work coauthored with several of his Berkeley graduate students, who conducted empirical analyses applying the extended case method (Burawoy et al., 1991). For them, the method is one of two ways to relate conditions in a given case (organization, neighborhood, social event) to the society at large in which it is embedded. In the interpretive method, Burawoy argues, the case reveals the essential nature of society at large, the way in Geertz's work the cockfight embodied the social organization of Balinese society (Geertz, 1973). In the extended case method, the case is understood by investigating the larger forces shaping conditions in the case – Geertz's cockfight would be understood as a 'ritual of resistance to colonial and then Javanese domination', and the economic and political forces shaping it would be the subject of analysis (Burawoy et al., 1991: 278; Burawoy, 1998). Burawoy's distinction seems to suggest that 'extending' is what the analyst does to understand the case: the analyst investigates society at large to determine its impacts on the case at hand.

But what is the basis of generalization in Burawoy's conception? By my reading, Burawoy is inconsistent on this point. On the one hand, he seems to suggest that the method is not inductive, in contrast, for example, to grounded theory. In fact, he defines the core of the method as one 'which examines how the social situation is shaped by external forces' (Burawoy et al., 1991: 6), which suggests that the ultimate purpose is to understand the social situation. As he explains when contrasting the method to grounded theory, while the latter develops 'micro foundations of a macrosociology', the extended case method develops the 'macro foundations of a microsociology' (Burawoy et al., 1991: 280; Burawoy, 1998), which seems to imply that the purpose is to understand the case, not to generalize from it.[17] On the other hand, Burawoy does provide at least part of an answer. Rather than 'statistical significance', he argues, the extended case method searches for 'societal significance. The importance of the single case lies in what it tells us about society as a whole rather than about the population of similar cases' (Burawoy et al., 1991: 281).

This appears to be an appropriate solution, until one wonders how it would work in practice – that is, what the ethnographer would do differently. One clue comes from understanding how the method would select cases, a practice that follows directly from van Velsen (1978 [1967]). Through the method, deviant or unique cases are especially interesting, because they provide for ways of developing or extending theories. This represents a potentially appropriate alternative to Bill's predicament.

But this line of thinking would suggest that the ultimate purpose is refining or reconstructing a theory (see Burawoy, 1998, 2003; Burawoy et al., 1991), rather than identifying an empirical fact about society, for which some understanding of the relationship between the case and other cases is important. For example, a theory might propose that small nonprofit organizations must be connected to stable funders to survive more than five years from birth in highly competitive environments. The extended case researcher might then find a small nonprofit that was not so connected but nevertheless survived 10 years, and ask why. Clearly, an interesting development of the *theory* might emerge from such a study. Less clear is what kind of empirical knowledge of *society* would emerge from it, unless one adopts an a priori model of society as a whole, which Burawoy and his students do from a perspective of 'domination and resistance' and 'globalization' (Burawoy et al., 1991, 2000). Once the neo-Marxist or globalization perspective (or a pre-determined alternative) is abandoned, the ethnographer is still at a loss as to the type of *empirical knowledge* gained.[18] Either it is knowledge about how one case works, or it is knowledge about how other cases work. If it is the latter, then some logical justification, some basis for feeling confident that another nonprofit with no connection to stable funders would exhibit similar patterns, is still missing.

In this sense, I suggest that Burawoy's (1998; Burawoy et al., 1991) extended case method is only a partial solution to Bill's particular predicament. It provides a potentially effective way of improving theories, by proposing (as many others have) the use of unique or deviant cases to improve on existing theories. It does not quite propose a model for distinguishing a good from a bad hypothesis, since this is not the way it orients its research. And it does not provide an explicit criterion to make empirical assessments relevant to other cases.[19]

Manchester revisited

An earlier proponent of the extended case method, Clyde Mitchell, provides a clearer answer to our particular set of questions. Mitchell's conception of the method elaborates on Gluckman's but differs from Burawoy's. In a 1983 essay, Mitchell, following Gluckman, distinguishes 'apt illustrations' from 'extended case studies' in that the latter require 'further elaboration of the basic study of case material' because they '[deal] with a sequence of events sometimes over quite a long period, where the same actors are involved in a series of situations' (Mitchell, 1983: 193). Whereas Burawoy believes the key to the method is explaining local conditions in light of external forces, Mitchell believes the key to the method is its ability to uncover process:

> The particular significance of the extended case study is that since it traces the events in which the same set of main actors in the case study are involved over a relatively long period, the processual aspect is given particular emphasis. The extended case study enables the analyst to trace how events chain on to one another and how therefore events are necessarily linked to one another through time. (1983: 194)

This focus is generally consistent with the idea that fieldwork should devote itself to uncovering mechanisms and tracing processes (Lamont and White, forthcoming). However, Mitchell does much more.

Mitchell takes up specifically the question Bill asked, in slightly different language: 'How do you know that the case you have chosen is typical?' (1983: 188). Mitchell's answer is consistent with mine: he argues 'that this question betrays a confusion between the procedures appropriate in making inferences from statistical data and those appropriate to the study of an idiosyncratic combination of elements or events which constitute a "case"' (Mitchell, 1983: 188). Mitchell believes that statistical representativeness is an irrelevant criterion, which implies that trying to find representative cases is a mistake. Thus, he argues, 'extrapolation is in fact based on the validity of the analysis rather than the representativeness of the events' (1983: 190).

The natural question is how to determine that an analysis is 'valid'. To answer, Mitchell contrasts 'statistical inference' from what he variously calls 'logical', 'causal', or (more impishly) 'scientific inference'.[20] The former is 'the process by which the analyst draws conclusions about the existence of two or more characteristics in some wider population from some sample of that population . . .'; the latter, 'the process by which the analyst draws conclusions about the essential linkage between two or more characteristics in terms of some explanatory schema' (1983: 199–200). Most quantitative research, he argues, employs both types of inference; case study methods can only employ the latter: 'inference from case studies . . . cannot be statistical and . . . extrapolability from any one case study to like situations in general is based only on logical inference' (1983: 200).

With a little elaboration, Mitchell's distinction between statistical and logical inference provides an effective way to distinguish good from bad hypotheses. Hypotheses based on case studies such as Bill's, Mitchell might argue, should never be statistical, only logical. Thus, it would be erroneous for Bill to hypothesize that because he observed a preponderance of, say, public cocaine consumption in his neighborhood then the average poor neighborhood will exhibit a preponderance of cocaine use in public. In this hypothesis, logic played no role; it is a descriptive inference based on one case, a mechanical connection between characteristics in two settings that would only have worked if the laws of probability could justify it. By this way of thinking, if Bill had observed a preponderance of cross-dressing, he would have been forced to hypothesize that the average poor neighborhood experiences a preponderance of cross-dressing. In fact, whether Bill's hypothesis 'sounds' right or not, the inference is both statistically unjustified and illogical – or more precisely logically unjustified, since there is no logical reason to believe it. This makes it an ineffective hypothesis.

But suppose Bill had observed that whenever a crime erupted in the neighborhood, some people retreated into their homes while others felt compelled to get organized. He also observed that those in the latter group had stronger connections to the neighborhood. Their parents had been raised there and their extended family lived nearby. Their mothers, fathers, and uncles had, years earlier, collaborated with other neighbors in the creation of the local community center and the foundation of a recreation area. Bill might have uncovered a causal relationship between attachment and participation. He might then hypothesize that the reaction to crime will depend on the strength of local attachment, such that those strongly attached (through various mechanisms) are likely to participate while those weakly attached are likely to retreat. This is not a descriptive hypothesis; it is a logical (or what Mitchell might call 'causal') one. While it still requires testing it is, nonetheless, logically justified, which makes it an effective hypothesis.

The former hypothesis took the form 'All entities of type A will exhibit characteristic Z', which was ineffective because there was no logical reason to believe it. The latter hypothesis takes the form 'When X occurs, whether Y will follow depends on W', which is logically justified given the processes observed. They are both hypotheses, to be sure, and both would require further testing. But only the latter is a good hypothesis (see Small, 2004).

One set of answers

Mitchell has brought us closer to solving Bill's predicament, by clarifying the role of logic in distinguishing appropriate from inappropriate hypotheses based on single-case studies.[21] Even if Bill only followed this principle his approach to his research would increase in sophistication. Notice that, for this model, whether Bill selected an ostensibly typical or an apparently atypical case would be *immaterial*, since the hypothesis would be based on a) logical inference and b) the internal conditions of the case.

While Mitchell provides a way to derive effective hypotheses, he still does not clarify what kind of *empirical* statement Bill might make. In this respect, Bill is limited but not powerless, since he can, with justification, offer one category of empirical statement – ontological statements, those regarding the discovery of something previously unknown to exist. That is, a well-executed single-case study can justifiably state that a particular process, phenomenon, mechanism, tendency, type, relationship, dynamic, or practice exists (Glaser and Strauss, 1967; Lofland and Lofland, 1995). This, in fact, remains one of the advantages of ethnographic work, the possibility of truly emergent knowledge.

Second alternative: sequential interviewing

When it comes to making empirical statements, Jane possesses an advantage over Bill. While we have already provided several ideas about how Jane might approach her research differently, there is a more radical solution to Jane's predicament. This solution involves rethinking not merely her objectives but also her understanding of each of her interviews. As I will show, this will yield not only better hypotheses but also a more grounded foundation to make empirical statements relevant to the cases not observed.

The foundation is to adopt Yin's (2002) distinction between *case study* logic and *sampling* logic (see Ragin and Becker, 1992). Yin's work is on case studies, and I am extrapolating from his work. I will argue that case study logic can be effectively applied to in-depth interview-based studies, such that the latter may be conceived as not small-*sample* studies but multiple-*case* studies. I do not believe all interview-based studies should be conducted in this fashion, but, for some objectives, it will likely produce the most effective answers.[22]

Sampling and case study logic approaches are different and fully independent ways of approaching data – they constitute wholly different languages. Sampling logic refers to the principles of selection associated with standard survey research. In a sampling model, the number of units (e.g. individuals) to be studied is predetermined; the sample is meant to be representative; all units should have equal (or known) probability of selection; and all units must be subject to exactly the same questionnaire. If conducted properly, the characteristics of the sample are expected to reflect, within a margin of error, those of the population as a whole. The objective is statistical representativeness. For example, a researcher designing a survey of political attitudes in the US would determine how many respondents she needs, develop a sampling frame, create a short questionnaire that every respondent is given, and ensure that her sample is statistically representative. The objective would be to make sure that the distribution of responses in her sample matched that in the population.

Case study logic, in Yin's terms, proceeds sequentially, such that each case provides an increasingly accurate understanding of the question at hand. In a case model, the number of units (cases) is unknown until the study is completed; the collection of units is, by design, not representative; each unit has its own probability of selection; and different units are subject to different questionnaires. The first unit or case yields a set of findings and a set of questions that inform the next case. If the study is conducted properly, the very last case examined will provide very little new or surprising information. The objective is saturation. An important component of case study design is that each subsequent case attempts to replicate the prior ones. Through 'literal replication' a similar case is found to determine whether the same mechanisms are in play; through 'theoretical replication' a case different according to the theory is found to determine whether the expected difference is found. Sampling logic is superior when asking descriptive questions about a population; case study logic is probably more effective when asking how or why questions about processes unknown before the start of the study.

To elaborate and demonstrate the fruitfulness of case-study logic, consider an example based not on interviews but on experiments. Alfonse conducts an experiment in which one group of black and white students at Cal Tech is told they will receive an IQ test and another is told nothing. Both complete the 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 that the fear of fulfilling a stereotype about low IQs among blacks is at play. This inference is based on this one case.[23]

However, Alfonse realizes that black and white students at Cal Tech might differ from those elsewhere, and that how college students take tests might differ from how other people take them. That is, he realizes that he did not have what statisticians would refer to as a representative sample of the population. In addition, he realizes that his theory is not, strictly speaking, about black people, but about the impact of fears of fulfilling stereotypes on performance. For this reason, he then conducts literal and theoretical replications. With a colleague at Duke, Alfonse repeats the experiment among Duke undergraduates (literal replication); back at Cal Tech, he repeats it, but using women and men instead of blacks and whites, since he knows that there are stereotypes about women's inability to perform well on math tests (theoretical replication). If the theory is right, Alfonse infers, it should work for anyone, not just blacks and whites.

Some tests 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 senior citizens, and on and on. With each experiment, he refines and re-evaluates his understanding of the process. Slowly, as the number of experiments (i.e. cases) increases, his confidence that his theory is right begins to grow. 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, even though he expected much more subtle relationships than he could have envisioned during the first experiment. At this point, he has attained saturation.

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 subsequent experiment were chosen deliberately, based on his increasingly refined and continuously re-evaluated understanding of the underlying phenomenon. Furthermore, he cannot make accurate descriptive statements about the distribution of characteristics in the population as a whole. Alfonse would not report that, for example, 80 percent of blacks are susceptible to stereotype threat, or that, since 75 percent of the experiments confirmed his theory, his theory is right 75 percent of the time (this would be wrong on many, many counts). Instead, he has identified, and confirmed beyond reasonable empirical doubt, an important mechanism affecting test performance. I suggest that this approach may be used to think about in-depth interview research.[24]

Reconsider Jane's predicament. Jane was applying, or attempting to apply, sampling logic, and by that logic she had a weak sample. By a case logic, she would have had a weak selection of cases as well, because *inferential logic* played little role in her selection of each subsequent interviewee. However, depending on her question, she might have attained saturation even if, in the end, the number of individuals interviewed was relatively small.

The key is to conceive of every individual the way Alfonse conceived of every single experiment, as a single case. Jane, without knowing how many respondents she will eventually interview, interviews one. After a two-hour interview, the person, say, recalls and recounts experiencing discrimination from Latino immigrants when she was a child, thus developing anti-immigrant sentiments and favoring Draconian immigration reform. From the interview, Jane has begun to form a new understanding of her initial issue, the attitudes of African Americans about immigration. She theorizes that native-born blacks who have experienced discrimination from Latino immigrants will favor immigration reform due to their personal experience. 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 include increasingly refined questions about different aspects of discrimination, since every interview is refining and forcing her to re-evaluate her understanding of the phenomenon. She repeats the process multiple times. Her last interviews are longer than the first, and they include many more subtle variations on the way one experiences discrimination, including many that had not occurred to her during the first interview. 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.[25]

Jane's new approach would violate 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. Consequently, Jane can make no accurate statement about the distribution of attitudes. She would not report that, say, 25 percent of working-class blacks favor immigration reform, just as Alfonse would not make statements about the distribution of his effect in the population of African Americans. However, we would have the same confidence in her empirical findings as we do in Alfonse's statements that stereotype threat reduces performance.[26]

As for the most common question, the answer is clear: she will know how many interviewees she needs when her study is over – that is, when she has attained saturation. This predicament is far less threatening than may appear, since Jane retains the power to determine how to refine and re-evaluate her theory. The more complex the process, the more interviewees she will need, and it is not difficult to imagine that, with the identification of a straightforward but unknown process or a complex but common one, Jane might attain saturation with the 35th carefully selected case. For a few years, my graduate students and I have been conducting a modified version of this type of sequential analysis in interviews with urban mothers whose children were enrolled in daycare centers in New York City. We were studying many aspects of their social networks (including several for which I was employing survey data), but one of the perplexing questions was how they developed so many friendships with other mothers, especially in a place where (I erroneously thought) all they did was drop off and retrieve their children. After no more than a few dozen interviews – with rich and poor mothers; with blacks, whites, and Latinas; in multiple boroughs; and with even a few men – my research assistants and I had attained saturation on the six or seven mechanisms through which mothers in New York City centers made new ties (the PTA, the yearly fundraiser, the mandatory meetings, etc.) (Small, forthcoming). I was relatively confident, and subsequent interviews confirmed, that we would not identify new mechanisms for friendship formation (or new opportunities for making friendships) that we had not already seen. On this particular question, we had early on attained saturation.[27]

Conclusions

For ethnographers in many fields of study, the issues discussed in this article do not arise. These ethnographers have no intellectual engagement with quantitative researchers, no expectation that the latter will ever be reviewers, no need to assess the work against a larger body of quantitative studies asking similar or related questions. For many others, however, the issues are critical. In urban sociology, the study of immigration, and social inequality – as well as, to some extent, the sociology of organizations and of education – ethnographers must contend, explicitly or implicitly, with scholars trained in radically different traditions who claim expertise on the same questions and may well assume a unity of method.

In this context, I have argued that no matter what they do within the parameters of their original projects, Jane and Bill will never build airplanes with the capacity to fly. The 'representative' single neighborhood does not exist.[28] Jane and Bill should not be building airplanes; they should be building, say, boats, vessels that are equally important for transportation and, during some circumstances, much more effective. And rather than build boats that try to fly, they should build boats that sail effectively. I have defended Mitchell's elaboration of the extended case method and proposed sequential interviewing – along with the (common) recommendations to consider sampling for range, snowballing, and identifying unique cases – as one set of tools, by no means the only ones, to build better vessels. Generally, the approaches call for logical rather than statistical inference, for case- rather than sample-based logic, for saturation rather than representation as the stated aims of research. The approaches produce more logically sensible hypotheses and more transparent types of empirical statements. Regardless of the method, ethnographers facing today's cross-methods discourse and critiques should pursue alternative epistemological assumptions better suited to their unique questions, rather than retreat toward models designed for statistical descriptive research. After all, to push the metaphor to its limits, even the best built airplane will never do well at sea, and I do not see any pilots losing sleep over it.

Notes

1 This article is a major elaboration of an earlier talk, 'Lost in Translation', prepared for a 2005 National Science Foundation workshop on interdisciplinary standards in qualitative research (Small, 2008a). I thank members of the workshop, as well as Stanley Lieberson, Scott Lynch, Sabrina Placeres, Loïc Wacquant, Yang Yang, the students in my logic of inquiry courses at Princeton University and the University of Chicago, and

the editors and anonymous reviewers of *Ethnography* for comments and criticisms.

2 *Social Problems* and *City and Community*, among the mainstream journals that publish papers in these fields, have exhibited stronger inclinations to publish ethnographic work.

3 In fact, the National Science Foundation recently felt compelled to organize two conferences, with accompanying reports, aimed at discussing what constitutes 'science' in the context of ethnographic or interview-based qualitative work (see Lamont and White, forthcoming; Ragin et al., 2004). The purpose was clear, to help qualitative researchers negotiate a predominantly quantitative discourse:

> Workshop participants were asked to: 1) provide guidance both to reviewers and investigators about the characteristics of strong qualitative research proposals and the criteria for evaluating projects in NSF's merit review process, and 2) provide recommendations to address the broader issue of how to strengthen qualitative methods in sociology and the social sciences in general. The workshop was intended to contribute to advancing the quality of qualitative research, and thus to advancing research capacity, tools, and infrastructure in the social sciences. (Ragin et al., 2004: 5)

4 The more standard practice to assess the work against other cases would resemble Sallaz's (2008) recent study, which examines not cockfights but casino card games in South Africa to assess the extent to which they also embody social power relations and reflect larger structural transformations.

5 To be clear, these and other manuals certainly address questions of research design, at length. However, they tend to say little explicitly about what might be the most important questions for quantitative researchers. In fact, the question of representativeness and statistical generalizability, probably the most sensitive and ambiguous in the discourse between quantitative and qualitative researchers, receives scant attention in the widely used Hammersley and Atkinson (1995; see pages 42–5). It receives even less coverage in Lofland and Lofland (1995).

6 Readers should note that they will not find in these pages an attack on quantitative methods (which I often employ in my work). At issue is the difference between an approach to science in which different questions are asked within different epistemological frameworks and one in which all social scientists attempt to answer all questions from one framework. It should become obvious why this article defends epistemological pluralism.

7 I think Lamont's study of 160 upper-middle-class men in France and the United States is a methodologically sophisticated interview-based study. 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 percent and 58 percent, by liberal estimates

(Lamont, 1992). (As Lamont [1992: 285] 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. Thus, the figures could overstate the response rate.) In addition, the samples are small, only 80 individuals in each country (40 in each site). 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; and her thoughtful selection of research sites (Lamont, 1992, Appendix II).

8 For a discussion of these issues from researchers aimed at bridging the qualitative/quantitative divide, see King et al. (1994, pp. 63ff).

9 A more accurate statement of her question is this: Jane would like to know that, were she to sample the same population repeatedly, 95 out of 100 times the true number of working-class blacks who are pro-immigration reform would fall within +/– five percentage points of the estimated value.

10 The formula is $n = (Z^2 \times p \times (1 - p)) / C^2$, 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 Internet, where one can manipulate the assumptions. For example, [www.raosoft.com/samplesize.html].

11 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' book. Rather, it is to show the pitfalls of this particular way of thinking about case study selection, which the authors share with many others. One could argue that the authors' flaw in this instance stems from their attempt to use a case study to find an 'estimator', in the quantitative sense. From that perspective, we may quote Campbell and Stanley, who said of 'the one-shot case study' that 'such studies have such a total absence of control as to be of almost no scientific value' (Campbell and Stanley, 1963: 6). I disagree strongly with Campbell and Stanley's implied assessment of the *purpose* of a case study; however, I agree that, if the purpose *were* to make a statistically generalizable statement of the effect of one variable on another, the authors would be right, and their critique would apply to King et al.'s 'case study estimator' (1994).

12 On the research dilemmas brought about by IRBs and the problems of confidentiality, see Shea (2000). Today, researchers have to make difficult decisions about how to report their findings, since IRBs often require ethnographers to guarantee the confidentiality of their respondents. All too often, the solution requires thinking of neighborhoods in the abstract, devoid of historical or political context, further reinforcing the tendency to think of cases within sample-based logic.

13 In fact, one of the advantages of this section of King et al.'s (1994) book, as seen through Bill's application, is that it takes the *logic* of many such studies – where the approach is *implicit* – to its natural conclusion. Some ethnographers in these fields do, in fact, select their neighborhoods or cases based on careful demographic assessments of whether the site is empirically average (e.g. Klinenberg, 2002; McDermott, 2006). Others select, and at times even stumble, onto their cases through more informal processes, but with a similar understanding that the site being studied is somehow typical (e.g. Pattillo, 1999; Venkatesh, 2000; Wacquant, 2007). (In fact, this is one of the strongest proclivities inherited from the Chicago School of sociology.) In both cases, the implication is that by studying the right neighborhood, the ethnographer has earned a greater warrant to speak of 'the ghetto', 'the American housing project', 'the immigrant enclave', or another equivalent type of case. Conversely (following that logic), had an apparently *a*typical neighborhood been studied – such as a housing project that managed to keep crime at bay – then the weight of the warrant would be radically lowered. Of course, some ethnographers are more careful than others in the care with which they subsequently make claims about other neighborhoods, or neighborhoods in the general 'type', based on the study of the 'typical' case. However, as we shall see, I argue that *no* single neighborhood, whether apparently typical or radically atypical, produces such a warrant; that there is little justification for studying apparently typical neighborhoods *within this context* and several motivations to explicitly study *a*typical ones; and that the language for discussing work in these contexts should change radically.

14 The median age for males and females is 36; 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; $36,911 is the median earnings for individuals for the year 2006. See Section 1, Population, of the *Statistical Abstract of the United States* [http://www.census.gov/prod/2007pubs/08statab/pop.pdf].

15 Early Chicago School criminologists adopted a version of this procedure in their book-length studies of a single juvenile delinquent (Shaw, 1930, 1931). In a discussion included in one of the studies, Burgess convincingly argues that an in-depth study of this nature is especially informative about

the processes of becoming a full-time offender, but then unconvincingly attempts to argue that the case is 'representative':

> The case of Stanley appears also to be typical in a more real sense than can be verified by any statistical calculation. It is typical (i.e., belonging to the type) in the same way that every case is representative of its kind or species. This case is a member of the *criminal species* . . . The individual person is more intrinsically a specimen of any group of which he is a member than is a plant or animal of its species . . . [T]he relation of the person to his group is organic and hence representative upon a cultural rather than upon a biological level. (Burgess, in Shaw, 1930, p. 186)

Burgess's use of metaphors here does little to address the fact that a completely different delinquent would have to be understood as equally 'representative', thus begging the initial question.

16 In his two articles proposing seven approaches from which to motivate ethnographies, Katz (2001a, 2002) suggests several additional perspectives. For example, when ethnographic data are praised as revealing either hidden or overlooked aspects of social life, discovery becomes more important than mere empirical applicability to other cases. (More on this below.)

17 This interpretation is reinforced by his global ethnography, where he argues that the larger functions that should play a role in understanding the case are not merely societal but global (Burawoy et al., 2000).

18 Interestingly, one might trace this tendency – of the method to rest foundationally on a theoretical proposition about society – to van Velsen (1978 [1967]). Van Velsen preferred the term 'situational analysis', because his concern was that, when studying a case, one should present it as a situation as a whole, with all of the local and external factors shaping the observed phenomena. When identifying the assumptions of his model, van Velsen (1978 [1967]: 146) writes: 'One of the assumptions on which situational analysis rests is that the norms of society do not constitute a constituent and coherent whole. On the contrary, they are often vaguely formulated and discrepant. . . . Situational analysis therefore lays stress on the study of norms in conflict.' In this respect, the method (situational analysis) presupposes a theoretical subject matter (norms in conflict), the way Burawoy's extended case method presupposed, or seemed to, a theoretical subject matter (domination and resistance).

19 Burawoy's later work would not so much solve the problem as change the question. In Burawoy (1998), Burawoy et al. (2000), and Burawoy (2003) he pushes, respectively, for a reflexive mode of science that takes the obtrusiveness of the researcher as a starting point; for an ethnographic program that aims to understand the local case in light of global conditions; and for an approach to revisiting sites in which not mere refutation but a revised

understanding of the role of structure in affecting both present and past conditions plays a role.

20 Many have made some version of this distinction: ‘statistical inference vs logical inference’, ‘statistical generalizability vs analytical generalizability’ (Yin, 2002), and ‘enumerative induction vs analytical induction’. This last distinction is Znaniecki’s (1934), whose predicament – defining case studies in an age inclined to rely on quantitative research to examine social problems – parallels our current one all too closely.

21 It is worth noting that many of the other, quite appropriate recommendations – that ethnographers engage in ‘process tracing’ or ‘unpacking mechanisms’ or ‘identifying conditions’ – are in essence more complex versions of this basic insight (see e.g. Lamont and White, forthcoming).

22 Some of the ideas that follow are consistent with grounded theory. In fact, the processes of literal and theoretical replication, described below, are similar to Glaser and Strauss’s (1967: 59) description of ‘theoretical sampling’. Readers will also notice that sequential interviewing has much in common with Bayesian inference. See also Katz (2001b).

23 The inspiration for this example is the work of Claude Steele and colleagues at Stanford University on ‘stereotype threat’ (see Aronson et al., 1999; Steele, 1992; Steele and Aronson, 1995). The example by no means attempts to represent the process through which Steele and his colleagues actually performed their research.

24 To be clear, I do not believe Alfonse can now make determinist statements. His understanding of the effect of stereotype threat would still fall within the parameters of a probabilistic understanding of the social world (Lieberson, 1991).

25 These descriptions of the research process are stylized, as all of them are forced to be. In real life, Jane might have interviewed 10 people before any semblance of a story had emerged. She then would have limited the scope of her study and her questions, to prevent continuing to interview indefinitely.

26 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.

27 A reviewer asked how this approach would work in the large, team ethnographic projects that have proliferated in recent years (e.g. Newman, 1999). In theory, there should be no difference in the application of sequential interviewing. However, the logistics of running such projects may render the careful application of a sequential process difficult, which happens to highlight one of their potential weaknesses. If one or two PIs supervise without entering the field, or if the team is especially large, or if the project takes place in multiple cities, it becomes difficult to coordinate early discoveries, to devise literal or theoretical propositions to replicate, or to shift direction collectively to answer newly pressing questions. Indeed, it is possible that as the number of ethnographers increases the more the project

will resemble a survey, rather than a study in which most of the discoveries are emergent.

28 One of my first-year graduate students, convinced of the logic of these arguments, nonetheless explained that 'it feels better' to pursue these pseudo-representative samples, since, irrational as it may be, she had been indoctrinated into worshiping the gods of statistical representativeness. To this I have no reply, except perhaps to paraphrase Goya: when reason goes to sleep, nothing but monsters comes to life.

References

Ainsworth-Darnell, J. and D. Downey (1998) 'Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance', *American Sociological Review* 63: 536–53.

Aries, E. and M. Seider (2005) 'The Interactive Relationship between Class Identity and the College Experience: The Case of Lower Income Students', *Qualitative Sociology* 28(4): 419–43.

Aronson, J., M.J. Lustina, C. Good and K. Keough (1999) 'When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat', *Journal of Experimental Social Psychology* 35: 29–46.

Burawoy, M. (1998) 'The Extended Case Method', *Sociological Theory* 16(1): 4–33.

Burawoy, M. (2003) 'Revisits: An Outline of a Theory of Reflexive Ethnography', *American Sociological Review* 68(5): 645–79.

Burawoy, M. (2005) '2004 ASA Presidential Address: For Public Sociology', *American Sociological Review* 70(1): 4–28.

Burawoy, M., J.A. Blum, S. George, Z. Gille, T. Gowan, L. Haney, M. Klawiter, S.H. Lopez, S.O. Riain and M. Thayer (2000) *Global Ethnography*. Berkeley: University of California Press.

Burawoy, M., A. Burton, A.A. Ferguson, K.J. Fox, J. Gamson, N. Gartrell, L. Hurst, C. Kurzman, L. Salzinger, J. Schiffman and S. Ui (1991) *Ethnography Unbound*. Berkeley: University of California Press.

Campbell, D.T. and J.C. Stanley (1963) *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin.

Cook, P.J. and J. Ludwig (1997) 'Weighing the 'Burden of "Acting White"': Are There Race Differences in Attitudes Toward Education?', *Journal of Policy Analysis and Management* 16(2): 256–78.

Dilthey, W. (1988) *Introduction to the Human Sciences*. Detroit, MI: Wayne State University Press.

Duneier, M. (1999) *Sidewalk*. New York: Farrar, Straus & Giroux.

England, P. and K. Edin (eds) (2007) *Unmarried Couples with Children*. New York: Russell Sage Foundation.

Evens, T.M. S. and D. Handelman (eds) (2006 [1983]) *The Manchester School*. New York: Berghahn Books.
Feynman, R.P. (1999) *The Pleasure of Finding Things Out*. Cambridge, MA: Perseus.
Fordham, S. and J. Ogbu (1986) 'Black Students' School Success: Coping with the 'Burden of "Acting White"', *Urban Review* 18(3): 176–206.
Frankfort-Nachmias, C. and D. Nachmias (2000) *Research Methods in the Social Sciences*, 6th edn. New York: Worth.
Garfinkel, H. (1967) *Studies in Ethnomethodology*. Cambridge: Polity Press.
Geertz, C. (1973) *The Interpretation of Cultures*. New York: Basic Books.
Glaser, B.G. and A.L. Strauss (1967) *The Discovery of Grounded Theory*. Hawthorne, NY: Aldine de Gruyter.
Gluckman, M. (1961) 'Ethnographic Data in British Social Anthropology', *Sociological Review* 9: 5–17.
Goering, J. and J.D. Feins (eds) (2003) *Choosing a Better Life?* Washington, DC: Urban Institute Press.
Hammersley, M. and P. Atkinson (1995) *Ethnography*, 2nd edn. London: Routledge.
Harris, M. (1974) *Cows, Pigs, Wars, & Witches*. New York: Random House.
Hermanowicz, J.C. (1998) 'The Presentation of Occupational Self in Science', *Qualitative Sociology* 21(2): 129–48.
Katz, J. (1997) 'Ethnography's Warrants', *Sociological Methods and Research* 25(4): 391–423.
Katz, J. (2001a) 'From How to Why: Luminous Description and Causal Inference in Ethnography, Part 1', *Ethnography* 2(4): 443–73.
Katz, J. (2001b) 'Analytic Induction', in N.J. Smelser and P.B. Baltes (eds) *International Encyclopedia of the Social and Behavioral Sciences*. London: Elsevier.
Katz, J. (2002) 'From How to Why: Luminous Description and Causal Inference in Ethnography, Part 2', *Ethnography* 3(1): 63–90.
King, G., R.O. Keohane and S. Verba (1994) *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
Klinenberg, E. (2002) *Heat Wave*. Chicago, IL: University of Chicago Press.
Lamont, M. (1992) *Money, Morals, & Manners*. Chicago, IL: University of Chicago Press.
Lamont, M. and P. White (eds) (forthcoming) *Report from Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Washington, DC: National Science Foundation.
Levitt, P. (2001) *The Transnational Villagers*. Berkeley: University of California Press.
Lieberson, S. (1991) 'Small *N*'s and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases', *Social Forces* 70(2): 307–20.

Lieberson, S. and F.B. Lynn (2002) 'Barking Up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science', *Annual Review of Sociology* 28: 1–19.

Lofland, J. and L.H. Lofland (1995) *Analyzing Social Settings*, 3rd edn. Belmont, CA: Wadsworth.

McDermott, M. (2006) *Working Class White*. Berkeley: University of California Press.

Mitchell, J.C. (1983) 'Case and Situation Analysis', *Sociological Review* 31(2): 187–211.

Morales, M.C. (2006) 'Review of *Villa Victoria: The Transformation of Social Capital in a Boston Barrio*, by Mario L. Small', *Contemporary Sociology* 35(3): 283–85.

Neckerman, K. (ed.) (2004) *Social Inequality*. New York: Russell Sage Foundation.

Newman, K. (1999) *No Shame in My Game*. New York: Knopf.

Newman, K. and R.P. Massengill (2006) 'The Texture of Hardship: Qualitative Research on Poverty, 1995–2005', *Annual Review of Sociology* 32: 423–46.

O'Connor, A., C. Tilly and L. Bobo (eds) (2001) *Urban Inequality*. New York: Russell Sage Foundation.

Pattillo, M. (1999) *Black Picket Fences*. Chicago, IL: University of Chicago Press.

Portes, A. and R. Rumbaut (2001) *Legacies*. Berkeley: University of California Press.

Portes, A., L.E. Guarnizo and W.J. Haller (2002) 'Transnational Entrepreneurs: An Alternative Form of Immigrant Economic Adaptation', *American Sociological Review* 67(2): 278–98.

Ragin, C.C. and H.S. Becker (eds) (1992) *What Is a Case?* Cambridge: Cambridge University Press.

Ragin, C.C., J. Nagel, and P. White (eds) (2004) *Workshop on Scientific Foundations of Qualitative Research*. Washington, DC: National Science Foundation.

Rumbaut, R. and A. Portes (2001) *Ethnicities*. Berkeley: University of California Press.

Saiedi, N. (1993) *The Birth of Social Theory*. Landham, MD: University Press of America.

Sallaz, J.J. (2008) 'Deep Plays: A Comparative Ethnography of Gambling Contests in Two Post-Colonies', *Ethnography* 9(1): 5–33.

Schudson, M. (1992) *Watergate in American Memory*. New York: Basic Books.

Shaw, C.R. (1930) *The Jack-Roller*. Chicago, IL: University of Chicago Press.

Shaw, C.R. (1931) *The Natural History of a Delinquent Career*. Chicago, IL: University of Chicago Press.

Shea, C. (2000) 'Don't Talk to the Humans', *Lingua Franca* 10(6): 26–34.

Small, M.L. (2004) *Villa Victoria*. Chicago, IL: University of Chicago Press.

Small, M.L. (2007) 'Is There Such a Thing as "The Ghetto"? The Perils of Assuming that the South Side of Chicago Represents Poor Black Neighborhoods', *City* 11(3): 413–21.

Small, M.L. (2008a) 'Lost in Translation: How Not to Make Qualitative Research More Scientific', in M. Lamont and P. White (eds) *Report from Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Washington, DC: National Science Foundation.

Small, M.L. (2008b) 'Four Reasons to Abandon the Idea of "The Ghetto"', *City and Community.*

Small, M.L. (forthcoming) *Unanticipated Gains*. New York: Oxford University Press.

Somers, M. (1994) 'The Narrative Constitution of Identity: A Relational Network Approach', *Theory and Society* 23(5): 605–49.

Steele, C.M. (1992) 'Race and the Schooling of Black Americans', *The Altantic Monthly*, April.

Steele, C.M. and J. Aronson (1995) 'Stereotype Threat and the Intellectual Test Performance of African Americans', *Journal of Personality and Social Psychology* 69(5): 797–811.

van Velsen, J. (1978 [1967]) 'The Extended-Case Method and Situational Analysis', in A.L. Epstein (ed.) *The Craft of Social Anthropology*, pp. 129–49. Delhi: Hindustan Publishing Corpn.

Venkatesh, S. (2000) *American Project*. Cambridge, MA: Harvard University Press.

Wacquant, L. (1997) 'Three Pernicious Premises in the Study of the American Ghetto', *International Journal of Urban and Regional Research* 20(2): 341–53.

Wacquant, L. (2007) *Urban Outcasts*. Cambridge: Polity Press.

Walzer, S. (1997) 'Contextualizing the Employment Decisions of New Mothers', *Qualitative Sociology* 20(2): 211–27.

Weiss, R. (1994) *Learning from Strangers*. New York: Free Press.

Wilson, W.J. (1996) *When Work Disappears: The World of the New Urban Poor*. New York: Knopf.

Wilson, W.J. et al. (1987) *Urban Poverty and Family Life Survey of Chicago, 1987* [computer file], 2nd ICPSR version. Chicago, IL: University of Chicago, Center for the Study of Urban Inequality/National Opinion Research Center [producers]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Worsley, P. (1968) *The Trumpet Shall Sound*. New York: Schocken Books.

Yin, R. (2002) *Case Study Research*. Thousand Oaks, CA: SAGE.

Znaniecki, F. (1934) *The Method of Sociology*. New York: Farrar & Rinehart.

■ **MARIO LUIS SMALL** is Associate Professor of Sociology and the College at the University of Chicago. His research focuses on urban poverty, neighborhoods, inequality and culture, and case study methods. He is the author of *Villa Victoria: The Transformation of Social Capital in a Boston Barrio* (2004), for which he received the C. Wright Mills Prize, and of articles published in journals such as the *American Journal of Sociology*, *Annual Review of Sociology*, *Social Forces*, *Theory and Society*, and *Social Science Quarterly*. His forthcoming book, *Unanticipated Gains: Origins of Network Inequality in Everyday Life*, employs qualitative and quantitative data on urban mothers to examine how routine organizations affect the advantages that people can secure from their networks. *Address*: 1126 E. 59th Street, Chicago, IL 60637, USA. [home.uchicago.edu/~mariosmall, email: mariosmall@uchicago.edu] ■