In the not-too-distant past, qualitative and quantitative researchers in sociology and other disciplines were embroiled in a protracted conflict sometimes referred to as “the paradigm wars.”¹ Field-workers accused their presumed opponents of conducting “positivistic,” unreflective, or simple-minded research; quantitative analysts chided theirs for doing “soft” or unrigorous work, or of writing “just-so” stories with little scientific backing. To the extent it was a war, quantitative research clearly held the upper hand, as economics, demography, statistical analyses, and public opinion surveys deeply shaped national policy decisions.² But quantitative researchers who ignored fieldwork did so at their peril, as a vast body of ethnographic and interview-based research had documented and offered important insight into the experiences of many of the populations that those national discussions were concerned with, such as school students, the unemployed, married couples, low-income families, employers, and immigrants.³ In all, the “wars” were a highly counterproductive conflict that made evident just how young, in historical terms, most social sciences are.⁴

But much has changed. While quantitative research arguably remains dominant in social science debates on important social
problems, over the past two decades qualitative scholarship has dramatically shaped how scientists, policy makers, and the public think about inequality, poverty, race and ethnicity, gender, education, health, organizations, immigration, neighborhoods, and families. It has helped us understand why neighborhoods matter; how schools shape children’s opportunities, expectations, and understanding of themselves; why people risk life and limb to cross the southern U.S. border; how employers evaluate potential hires; how people make decisions about marriage, employment, and child-rearing; how people manage legal institutions; how social and economic conditions affect people’s everyday lives; and much more.

In contrast to the past, economists, demographers, political scientists, and quantitative sociologists of all stripes today often cite qualitative studies in their own work and use those studies to generate hypotheses, illustrate discoveries, or interpret findings. In turn, qualitative researchers have testified before Congress, helped governments set policy, advised local practitioners, contributed to the public discourse, and shaped how corporate and nonprofit boards invest and spend their resources. At this juncture, the importance of interview and ethnographic methods to social science, and to society, is not in question.

But in spite of this progress, social scientists have not come to agree on what constitutes good qualitative social science—in fact,
they do not even agree on whether qualitative research should be thought of as scientific, as opposed to merely informative, work.\textsuperscript{15}

How to Assess Quality

Consider the recent history. In 1994, Gary King, Robert Keohane, and Sydney Verba’s *Designing Social Inquiry* was supposed to lay to rest questions about the scientific foundations of qualitative studies by presenting clear guidelines for conducting and evaluating such work based on basic principles that guide quantitative research, such as reliability, unbiasedness, and efficiency.\textsuperscript{16} The book was unusually detailed, comprehensive, and full of examples, promising to unite qualitative and quantitative researchers under a common view of rigorous empirical science. Instead, it sparked even more controversy.\textsuperscript{17} To this day, the work strongly divides researchers, and qualitative researchers have repeatedly complained that its guidelines are inappropriate.\textsuperscript{18}

In the early 2000s, the National Science Foundation (NSF), which funds social science scholarship, was facing an “increasing number of [submissions for] qualitative research projects” in sociology and recognized both that many reviewers did not know how to evaluate the work and that those who did seemed to disagree on the appropriate criteria.\textsuperscript{19} The NSF convened a team of sociologists and later another of social scientists across several disciplines to clarify the standards by which qualitative research should be deemed rigorous and the criteria against which reviewers should evaluate proposals.\textsuperscript{20} Distinguished social scientists participated in both teams, and each working group produced a volume of papers with summary guidelines.
But the researchers had so little in common in approach and perspective that many of their summary recommendations merely restated basic principles one would expect any proposal would have, rather than offering many new criteria specific to qualitative work. For example, the set of standards released by the first team recommended that researchers should “write clearly and engagingly for a broad audience,” “locate the research in the relevant literature,” “provide evidence of the project’s feasibility,” and so forth. The second team had similar ideas, such as “situate the research in appropriate literature,” “pay attention to alternative explanations,” and “specify the limitations of the research and anticipate potential reviewer objections.” All of these recommendations have value, but they are suggestions that most experienced researchers in any discipline would already know to follow, rather than guidelines either specifically relevant or new to field-workers.

Not all of the NSF recommendations had this character; some were in fact distinctively appropriate to qualitative research. For example, the volumes recommended that researchers “assess the possible impact of the researcher’s presence [and] biography”; discuss the researcher’s “cultural fluency, language skill, . . . knowledge of particular research context”; and “describe and explain case selection.” But the guidelines nonetheless sparked controversy: at least one team member, who is one of the most distinguished and influential ethnographers in American history, openly challenged the NSFs conceptions of good research and produced an extraordinary minority report. In the end, the NSF efforts, important though they were, did not bring closure to the questions the experts had been tasked to address.

In fact, over the ensuing years this ongoing ambiguity has contributed to several high-profile disputes involving fieldwork.
Famous ethnographers with widely discussed books have been accused of handling data poorly, of making implausible claims, and of rendering their findings unverifiable.\textsuperscript{25} A recent book has “interrogated ethnography,” proposing that claims in several major ethnographic studies do not stand up to legal standards of verification.\textsuperscript{26} Indeed, some of the most contentious reviews of qualitative research today are produced by other qualitative researchers. Among the long list of controversies are those over grounded theory, “cowboy ethnography,” replicability in recent ethnographies, exaggeration in anthropology, sampling in case studies, the value of interviews in studying behavior, and the ethics of masking identities in qualitative research.\textsuperscript{27}

These controversies have left budding field-workers uncertain about how to conduct their own work; reviewers unclear about what signs of quality to look for; and scholars, journalists, and other consumers unsure about how to judge the work that qualitative researchers are generating.

Improving Social Science

That general uncertainty exists in a societal context where social science as a whole is looking more closely at its methods. In recent years, quantitative social science has experienced a reckoning, as common research practices in psychology, economics, political science, and quantitative sociology have come under extensive scrutiny. Critics have pointed to flaws large and small. Some problems are serious but relatively benign, faults of omission rather than commission; for example, studies in many subfields have been shown to be all too dependent on “WEIRD” samples, from white, educated, industrialized, rich, democratic respondents, a dependency
that calls into question the work’s empirical generalizability.\textsuperscript{28} Other problems lie at the heart of scientific practice itself. Authors of high-profile quantitative studies have been found guilty of p-hacking (mishandling data to produce significant findings), HARKing (writing up hypotheses after the results are known), and even fabricating results.\textsuperscript{29} In fact, the findings of numerous major experimental studies have been uncovered to be nonreplicable.\textsuperscript{30} These practices undermine not only the quality of the science but also the public’s trust, a problem that, in an era in which some politicians have sought to question the reliability of science, can be especially pernicious.\textsuperscript{31}

Faced with this reckoning, however, quantitative researchers have instituted solutions. Journals have formalized many practices to improve how transparent and accountable quantitative studies are, including publishing more papers that aim to replicate prior findings, encouraging researchers to preregister their hypotheses in public repositories, requiring authors to publicly post the data and code that produced their analyses, and more.\textsuperscript{32} At least some of these efforts are working. Malfeasance, mistakes, and questionable analytical decisions are often discovered quickly, discussed openly, and retracted as needed.\textsuperscript{33} Bad research habits that were commonplace a decade ago are increasingly rare in the disciplines’ best journals, as editors impose greater transparency, expect multiple robustness checks, and require making code and data easily available to reviewers.

These developments might have been great news for qualitative researchers but for the fact that many of the recommended practices are inappropriate for their work. For example, calling for researchers to replicate more ethnographies will often make no sense: an observational case study of a single event—for example, the Arab Spring or the George Floyd protests—cannot be replicated,
since a future researcher cannot return to the past, recreate the event, and experience what the real-time observer did. Similarly, requiring all field-workers to preregister their hypotheses would undermine major traditions in which the research is inductive; an ethnographer who is entering a study site without intending to test a prior hypothesis would have nothing to preregister.

In recent years, some qualitative researchers have proposed alternative ways of increasing the transparency of qualitative research. Some scholars have discouraged ethnographers from anonymizing a field site unless necessary. Others have called for interviewers to make de-identified transcripts publicly available. But these proposals are not themselves universally agreed upon. Moreover, they will be inappropriate for many kinds of projects, such as those in which identification poses too great a risk to participants or in which complete de-identification would render a set of transcripts all but meaningless. Finally, while increasing transparency will improve some kinds of work, it cannot be the heart of the solution to our problem, for it does not inherently distinguish good from bad work. Consider the proposal to resist anonymizing field sites: many weak ethnographies are open about where they were conducted, while in contrast, one of the most important ethnographies in history, William Foote Whyte’s *Street Corner Society*, anonymized its site by calling it “Cornerville.”

While transparency might improve accountability, it still does not address our motivating question: how to assess the soundness of qualitative work. We have posed this question—what criteria would you use to distinguish empirically sound from unsound qualitative social science?—to many researchers across methodological perspectives and disciplines, including not only sociologists but also economists, psychologists, and statisticians, and across career
stages, from senior scholars to graduate students beginning their fieldwork. Many have confessed that they ultimately do not know how they would answer. Our objective is to present a nonexclusive set of criteria applicable to any social scientist conducting in-depth interviews or participant observation. Our criteria will likely be recognizable by any qualitative researcher who conducts field-based research, since, as we shall see later, and despite their public epistemological debates, field-workers often demonstrate tacit agreement about quality in craftsmanship. We begin by clarifying some terms.

What Is “Qualitative Research”?

In a strict sense, there is no such thing as “qualitative research.” There is no single research practice, perspective, attitude about data, or approach to social science that all scholars who have used that term to describe their work follow. As one of us has written,

Some use the term “qualitative” to describe all small-sample studies, regardless of whether the analysis is formal, because they consider those studies to lack statistical generalizability. Others use “qualitative” to characterize any approach in which units (such as organizations or nations), regardless of their number, are analyzed as cases rather than divided into variables, such as studies of revolutions in which countries are assessed in light of their particular historical circumstances. Still others use the term to refer only to studies that rely on hermeneutic or interpretive, rather than positivistic orientations. Because of these differences, the quantitative versus qualitative opposition has been used to contrast many kinds of alternative studies: large-n versus small-n, nomothetic versus idiographic, causal versus interpretive, variable-based versus...
case-based, explanatory versus descriptive, probabilistic versus deterministic, and numerous others.\textsuperscript{41}

One way of making sense of this heterogeneity is to understand that “qualitative” can refer to at least one of three different elements of the research project: the type of data, the method of data collection, or the approach to data analysis. “The first refers to that which has been collected for study; examples are interview transcripts, . . . newspaper clippings, field notes, and administrative records. The second refers to the means for obtaining data,” such as interviewing or participant observation.\textsuperscript{42} “The third refers to the means for making sense of the collected data.”\textsuperscript{43} To analyze the data, a researcher can choose among a broad range of both specific techniques, such as open versus closed coding of interview transcripts, and general approaches, such as grounded theorizing versus the extended case method.\textsuperscript{44}

No type of data or method of data collection requires any specific analysis technique. One can use interview transcripts to write extended biographical narratives or to run statistical regressions.\textsuperscript{45} One can use observational field notes to write classic ethnographies or to produce formal quantitative analyses.\textsuperscript{46} Thus, there is little sense in trying to propose a single definition of “qualitative research.” Our book uses the term merely as shorthand.

\section*{Our Focus}

\textit{Methods of Data Collection}

Still, our focus is narrow. Among the three elements—types of data, method of data collection, and type of analysis—our main