FOUR

MATERIALIZING CLEAN DATA IN THE FIELD

It is early morning in mid-February 2008, and the MAYP SUV forges its way awkwardly through grasses taller than its roof, becoming mired every few kilometers in mud. Each time its wheels spin in the muck, we exit the vehicle to help dislodge it. Before fieldwork begins for the day, we are thoroughly covered in mud. Anticipating a long day with many impediments to smooth data collection, the eight of us (a driver, a supervisor, five interviewers, and myself) packed a small cooler with bottles of water, yogurt, and Mahewu (a grainy maize drink that is a favored field lunch). We also carry a loaf of bread, knowing it will be difficult to find chips stands or tearooms in the remote enumeration area (EA) that is the site of today's fieldwork.[1] The EA is located in Thuma Forest Reserve, an area of rugged topography in central Malawi about twenty kilometers from the main road connecting the capital, Lilongwe, to the lakeside town of Salima. Each of the five interviewers will visit three households by the end of the day to collect survey data for MAYP. Spirits are a bit low on the heels of a frustrating few days of fieldwork dogged by flooded bridges, impassable roadways, long walks in water-saturated shoes,

FIGURE 4.1. An SUV belonging to MAYP stuck in the mud, 2008. Photo by the author.

and the slow progress common to fieldwork during Malawi's rainy season. Pushing the SUV and slipping in the mud, fieldworkers recall other rainy season fieldwork mishaps, laughing about the time they hired canoes from local people and navigated through "crocodile-infested waters" to visit sample households unreachable by a washed-out road (see figure 4.1).

As we slowly make our way toward the EA, Chifundo, the team supervisor, opens a thick brown folder with the EA's number written on it in black marker and distributes to each interviewer a collection of items: three questionnaires, consent forms, crude maps of the area drawn by teams in previous years, headshot photos of assigned respondents (referred to as "snaps"), bars of soap for gifts, and yellow handheld GPS devices to be programmed with household coordinates. We scrutinize the maps to plan a time-efficient strategy of attack, and the SUV stops frequently to allow interviewers to disembark one by one, sometimes still a few kilometers' walk from their assigned households. Most interviewers carry umbrellas to cope with intermittent downpours. Chifundo points to a baobab tree that rises above grasses that stretch as far as the eye

can see, indicating that the SUV will wait at this landmark to collect all the interviewers at the end of the day. One interviewer returns to the tree shortly after being dropped off, unable to locate his assigned household. Chifundo sets off in search of the local chief to inquire about its location, meeting two men in army fatigues who patrol the reserve for poachers.

As interviewers finish their assigned interviews, they return one by one, covered in mud, to the SUV to submit their completed questionnaires to Chifundo and myself for checking, play bao with the curious young children who congregate near the SUV, sleep, or listen to music. By the end of the day, thirteen of fifteen damp questionnaires are successfully filled in. The team groans in frustration: we will have to return to the bush again in the coming days to find and interview the two respondents who were not at home today (a man who was out buying maize and a woman who was at the district hospital delivering a baby), consuming time and fuel in the process. Chifundo takes this news ambivalently: "These are the challenges we face *kukapita field* [going off to the field]!"

. . .

This scene, re-created from my field notes, foregrounds the logistical challenges faced by fieldwork teams, especially on rainy days when data are being collected in remote areas like Thuma. The SUV caught in the mud is a fitting metaphor for the messy impediments projects like MAYP encounter everyday in their quest to collect clean data. While the ideal vision of researchers conjures efficient interviewers visiting all sample households and recording accurate data as neat pencil marks on questionnaires, fieldwork teams find themselves navigating many unexpected obstacles in the field. Distant from the eyes and ears of the demographers and economists who design the surveys and outfit teams with maps, clipboards, and other accoutrements meant to streamline data collection, fieldworkers embody—if imperfectly—the epistemological investments of their employers. Fieldwork places a set of demands on perception, subjectivity, and performances that help materialize data. Nonetheless, tensions between the abstract standards that govern data collection and the material circumstances of the field engender creative tactics on the part of fieldworkers who seek to manage, if not eradicate, uncertainty and errors in the data they collect.

As will become clear in this chapter, collecting clean, high-quality data entails learning to "see like a research project" (Biruk 2012). Not unlike James Scott's (1998) state, survey projects in Malawi utilize tools and technologies to better see their subjects: maps, questionnaires, photos, GPS devices, and

sampling, for example. These tools collect and organize heterogeneous information that is converted into valuable numbers and are central props in structuring ways of seeing, gestures, and other forms of body work exhibited by fieldworkers (Boyer 2005, 259–260; Vertesi 2012). As a supervisor told a new crop of LSAM data collectors during a prefieldwork training session, "*You* are the project."

In what follows, I trace how researchers' scientific investments in pure, clean data—symbolically represented in surveys that act as a recipe for data collection—are made and unmade by practices and processes on the ground. Through close analysis of the embodied techniques and technologies employed by fieldworkers during data collection, I illustrate how frictions between epistemological metrics for data and the particularities of everyday fieldwork produce—and come to validate—the numerical evidence we use to understand the AIDS epidemic in Malawi. I focus, in particular, on the cultural translation of survey concepts such as probability, the techniques and technologies used by fieldworkers to uncover the truth of rural Malawian social realities, and researchers' intensive efforts to harmonize encounters between fieldworkers and research participants. The chapter pays careful attention to how evidence is fashioned through technologies and relations that add value to numbers and codes recorded on a page, even as those processes also threaten to undo that value by cooking them, in the eyes of project designers.

In highlighting the production of data's value within the social relations and processes that make up the fieldwork phase of research, I bring to light the provisional and contextual nature of the value and uses of quantitative evidence that we usually encounter in a form detached from its contexts of production (Guyer et al. 2010; Lampland 2010; Ballestero 2012; Erikson 2012; Sangaramoorthy and Benton 2012; Day, Lury, and Wakeford 2014). Chapter 2 shows how fieldworkers perform and cultivate a marketable kind of local expertise aligned with researchers' expectations and described how data collection relies on the production of a spatiotemporal difference and distance between the field and the office. This chapter likewise centers fieldworkers' role in assembling data, but presents a fine-grained analysis of the nature of their interactions with data themselves; it considers how their bodies, affects, and practices in the field and the data they collect are coproduced. We will see that the embodiment of standards for clean data by fieldworkers is a central part of the coordination of data collection across thousands of research encounters.

As elaborated in chapter 1, the material form of the survey questionnaire, with its text waiting to be read aloud to respondents, boxes waiting to be

checked, and empty space waiting to be filled in by data collectors, is a template for the collection of good data in the field. The questionnaire plays a key role in the inscription processes of survey fieldwork by acting as a script for interviewers who are meant to translate the heterogeneous realities they document into usable units of data as they record them on the page (Callon 1986; Latour 1987). By shared demographic standards, data are expected to be clean: accurate and reliable, efficient and timely, and collected from sufficiently large, pure, and representative samples. The visions of researchers produce and rely upon conventions and tools that are organized, but not governed or controlled, by any one actor, and both enable and limit the movements and perspectives of those who populate research infrastructure (Knorr-Cetina 1999, 11). A survey project's fieldworkers need not visit every household in a given village to administer surveys, but only those included in the project's predetermined sample, for example. Researchers' investment in the sample as reservoir of data trickles down to fieldworkers whose everyday movements and interactions become conduits through which abstract disciplinary values and designs are translated into the field. The questionnaires they administer are boundary objects, a means of translating between intersecting social worlds (the village, the research project, the office, and policy), and various social groups (villagers, interviewers, data entry clerks, researchers) (Star and Griesemer 1989).

From start (survey design) to finish (eventual publication of articles based on survey data), the assembly line envisioned by researchers confronts threats, many of which arise during data collection in the field: mistranslation, lying respondents, respondents who refuse to participate, respondents who have migrated or are out of town, interviewer effects, poor weather conditions, inaccurate data entry, and lost data. High-quality, clean data attain value from their relative scarcity: not all projects can equally invest the resources, time, and energy needed to effectively manage uncertainty, as defined by a set of demographic epistemological norms. Fieldwork is expensive: fieldworker salaries, per diems, lodging costs, fuel, and constant car repair are some of the expenses evident in the opening scene of this chapter.

Scholars, institutes, and policy makers seek out data whose brand they trust and are familiar with; numbers and statistics carry the aura of the research project that produced and packaged them. Andrews, a longtime fieldwork supervisor with LSAM, reflected on the difference in brand between data collected by the June 2008 Malawi National Census and the data being collected by LSAM at the same time: "Those guys [National Statistical Office, NSO] are just hiring whoever because they need so many people to enumerate.

This is bad—their data will have problems. You can just look back to 1998 [year of the last census] to see how many problems come up with the data, all from hiring people [fieldworkers] without experience!" Andrews's endorsement of the LSAM brand devalues NSO data as flawed or dirty. High-quality, clean data are a vestige of a distant local reality faithfully and authentically captured by experienced and trustworthy fieldworkers and arbitrated at all steps along the way by checks and audits (Lyberg and Biemer 2008, 421).

Taking demographers' epistemic investment in high-quality, clean data as an entry point, this chapter argues that seeing like a research project necessitates standardization of habits, scripts, practices, and social interactions across thousands of social encounters in the field. It also shows how the unfolding practices and instruments of fieldwork shape the very objects they are meant to count and track (Haraway 1989, 171–172; Mol 2002; Asdal 2008; Lorway and Khan 2014). As Kapil Raj (2007, 226) suggests, the stabilization and collection of immutable units of information by fieldworkers associated with the nineteenth-century Indo-British exploration of Central Asia was rooted in the mutable nature of men themselves, and the knowledge and skills they embodied. Jamie Lorimer (2008, 391), too, highlights how surveyors for the U.K. Corncrake Census learned to reorganize their bodies and senses to better see, hear, and count corncrakes, a species of migratory bird. Yet because the standardizing values of enumerative projects are materialized in fieldworkers' bodily techniques (Mauss 1973), they also enfold uncertainty, which manifests in numbers that are profoundly provisional, even as they are immensely valuable as expedient placeholders for realities (Lampland 2009; Verran 2013). Standards of data collection make stability and fixity in numerical representation possible, despite—or perhaps because of—their customization by fieldworkers in the field.

Clean Data, Messy Field

> The completed questionnaire must be NEAT, CLEAR, READABLE, ACCURATE, UNBEND [*sic*], AND CREASE OR OIL FREE. . . . The questionnaires you are using are very sensitive to any manhandling. They should be kept unsoiled.
> —2008 Population and Housing Census Enumerator's Manual (NSO, Zomba, Malawi)

The mandate for clean, unsoiled questionnaires is taken from manuals distributed to enumerators for the Malawi National Census in 2008; it invokes the tension between clean and dirty data that likewise preoccupied LSAM, GSIP, and MAYP in 2007–2008. The imperative delivered from NSO to a cohort of

enumerators demonstrates an explicit aversion to bent, creased, oily, and messily written questionnaires in their material, paper form, but, more importantly, it draws a link between the questionnaires' physical forms and the quality of the data they will produce. During the 2008 census exercises, in fact, enumerators complained that they needed raincoats and other materials to protect census documents from winter rains and warned the NSO that if they were not properly equipped, data would be lost (Phiri 2008). Similarly, the district commissioner of Kota Kota (present-day Nkhotakota) in 1939 was concerned that census sheets distributed to village headmen to track basic demographics in their villages were—in the absence of a binder or container in which to collate them—so "dirty, dog eared and torn" as to be completely illegible (CAA 1939). Unsoiled questionnaires are the initial step in producing clean data, and maintaining the purity of the survey's white paper in the face of dust, rain, and greasy fingerprints is a fitting metaphor for the labor that goes into making clean data. In this section, I illustrate how clean data—usually considered to be an after-the-fact product of statistically based data cleaning or scrubbing procedures in the office—are an epistemic commitment that places demands on fieldworkers' perceptions, practices, and bodies in the field. Data and their collectors are made and remade by one another as data are assembled.

In order for them to achieve value for audiences who seek to use them, data must be accurate and reliable. Accuracy dictates that data must be as true a representation of reality, an individual, or a social phenomenon as possible. Reliability mandates that data and findings resulting from them must be replicable—obtainable in the same form again and again. Data cleaning is typically a method of dealing with data problems that occur: it can be glossed as the screening, diagnosis, and treatment of suspected errors in compiled data. Finding such errors requires familiarity with all phases of data flow, as errors can arise from bad initial planning, inadequate piloting (of surveys and people), and so on (van den Broeck et al. 2005). Common sources of error include missing data, input errors by data entry clerks, fabricated or invented data, coding errors, and interviewer or measurement error.[2] Though data cleansing or scrubbing techniques are usually applied to data that are already housed in databases, my informants emphasized the importance, as well, of keeping data clean during fieldwork.

Dirty data, from fieldworkers' perspectives on the ground, implied spelling mistakes or wrong numerical codes, forged or cooked data, incorrect data associated with a question, incomplete or sloppily entered data, missing data, or duplicate data. Field teams were well aware of their role in the larger process of

making clean, valuable data. If an interviewer neglected to ask a question of a respondent, for example, the blank space on the survey page became a stumbling block later on for the data entry team member who must enter that blank space into the database as "missing data." (During a training session for LSAM, a supervisor, Esau, informed a new crop of interviewers, "The absolute worst crime you can commit is 'missing data.'") Collecting data that are accurate and reliable entails meticulous attention to both linguistic and cultural dimensions of translation and to harmonizing and surveilling the behaviors of interviewers and data entry teams in prefieldwork training sessions and the field.

THE PROMISE AND PERILS OF BEANS: VERNACULAR PROBABILITIES

In chapter 1, I discuss that a major objective of prefieldwork survey design sessions and meetings between foreign and Malawian researchers is to translate hundreds of survey questions from English into local languages—Chewa, Yao, and Tumbuka—and to anticipate how such questions might be confusing to either respondents or interviewers. In addition to linguistic translation, survey design and fine-tuning necessitated attention to what might be termed accurate cultural translation. The twenty-five-page survey used by LSAM consisted of nineteen sections ranging from "Group Membership and Social Capital" to "AIDS," to "Marriage," to "Economic Situation," and so on. One of these sections, titled "Expectations Questions," assessed respondents' subjective expectations of future outcomes such as HIV infection, economic shocks, or illness. Researchers suggest that understanding such expectations is crucial to designing and evaluating policies in health, education, and so on (Attanasio 2009; Delavande, Giné, and McKenzie 2011).

This section of the LSAM survey was identified as a problem by interviewers and supervisors, making it an ideal site for exploring the potential and pitfalls of translating potentially complex concepts (here, probability) into simplified forms for a target audience with low literacy. In an attempt to ensure clarity of meaning of probability for its low-literacy sample of rural Malawians, LSAM implemented an exercise using beans that came to be known as *nyembanyemba* (beans, reduplicated) among fieldwork teams and research participants. Respondents were asked to place a certain number of beans in a dish to estimate how likely it was that they would, for instance, experience a food shortage or contract HIV/AIDS (one bean if it was unlikely to happen, ten beans if it was certain to happen; see figure 4.2). As an interactive elicitation technique, researchers consider the beans to be visual, intuitive, and fairly engaging for respondents and, importantly, view it as a translative

X2 Pick the number of beans that reflects how likely you think it is that…	# of beans in plate
a) You will have to rely on family members for financial assistance in the next 12 months	[_____]
b) You are infected with HIV/AIDS now	[_____]
FOR MARRIED RESPONDENTS **(INTERVIEWER: if respondent is not married → X2f)**	
c) Your spouse is infected with HIV/AIDS now	[_____]
FOR UNMARRIED RESPONDENTS	
d) Your romantic partner is infected with HIV/AIDS now (INTERVIEWER: if no romantic partner, write 99 and → X2h)	[_____]
e) You will be married one year from now	[_____]

FOR BOTH MARRIED AND UNMARRIED RESPONDENTS	
X3 Consider a healthy woman in your village who currently does not have HIV. Pick the number of beans that reflects how likely you think it is that she will become infected with HIV…	# of beans in plate
a) During a single intercourse without a condom with someone who has HIV/AIDS	[_____]
b) Within the next 12 months (with normal sexual behavior)	[_____]
c) Within the next 12 months if she is married to someone who is infected with HIV/AIDS	[_____]
d) Within the next 12 months if she has several sexual partners in addition to her spouse	[_____]
e) What about if this woman we just spoke about [in X3d] uses a condom with all extra-marital partners? How many beans would you leave on the plate?	[_____]

FIGURE 4.2. The beans exercise from the LSAM questionnaire, 2008.

technology that promises to increase quality and value of data collected from an imagined villager (Delavande and Kohler 2007; Delavande, Giné, and McKenzie 2011).

Respondents' and fieldworkers' responses to the beans were largely negative. Research participants tended to view the beans as infantilizing (a common reaction was, "If you want to play, go over there with the children!"), and the beans were an important site of friction between actors across different levels of the project.[3] Fieldwork supervisors negotiated carefully between top-down efforts to standardize implementation of this activity, their own skepticism about the beans, and the incessant complaints from fieldworkers that the beans exercise was silly, time consuming, and boring for respondents. Supervisors chastised interviewers for being lazy and encouraged them: "Improve your attitudes—the bad morale among your villagers [research participants] is coming from you! These guys [respondents] observe us. They can tell you think nyembanyemba is *chabe* [worthless] and this allows them to protest [against it]." They also occasionally spied on interviewers as they interviewed respondents to ensure they were not cheating the project by failing to do nyembanyemba and just filling in numbers at random (the most flagrant form of cooking data) in the boxes provided in the beans section. However, at nightly meetings with American researchers, the supervisors suggested that the beans exercise was a "misfit with Malawian culture" and difficult for Malawians to understand. They also suggested that respondents grew bored with the instrument and observed that they "tended to pick the number you give as example" when demonstrating the exercise. For example, he explained, if you taught the respondent about the beans using five beans as a halfway point between a high chance of rain today and no chance of rain today, respondents tended to continue to pick five throughout the remainder of the exercise.

A culturally relevant tool from the perspective of the researchers was, in local estimation, a failure in the Malawian cultural context.[4] At a technical level, fieldworkers complained that respondents often suggested they couldn't know what would happen in the future or suggested that only God could know such things. We note that nyembanyemba, a script for the capture of individual datums that would later become evidence of the probabilistic orientations of rural Malawians, became a site of struggle where data were malleable entities, perhaps more representative of negotiated research encounters than the rural reality they sought to represent.

My field notes recorded at households where nyembanyemba was implemented highlight some of the issues that arose when this tool was translated

into the field and make clear that numbers recorded in the boxes in the beans section of the survey are contingent and still unsettled renderings of the realities they seek to enumerate. In June 2008, Tapika, a twenty-four-year-old female LSAM interviewer, interviewed a thirty-five-year-old man, Josiah, in a village in central Malawi. After he showed us the tobacco balers he purchased to collect fees from his fellow villagers who used it, the pair (and I) sat behind his house on a mat he set out on the ground, and the survey interview proceeded smoothly until we reached the beans exercise (Section 15). Although Josiah was initially a bit baffled by the instructions ("I really should do this? [Move the beans around.] Can't I just answer the questions?"), he was a relatively willing participant.[5] Halfway through the long section, however, he grew tired of the beans and began to mention numbers without manipulating the beans and the dish in front of him. At this point, Tapika grew visibly frustrated and proceeded to pick up the number of beans Josiah said each time and place them in the dish, as if to indicate that Josiah must continue to use the beans. Josiah grew increasingly annoyed, and the defeated Tapika completed Section 15 without the beans.

In this encounter, Josiah made known his own reasonableness by making an effort to go along with the beans exercise he initially found unappealing. His later lack of interest, however, marked his effort to disengage from a social dynamic in which an interviewer asserted her status by requiring him to play with the beans. Tapika, as a younger woman interviewing an older man, negotiated the relationship carefully and likely felt compelled to perform the scripts and standardized implementation of the beans she had learned in training sessions, not least for the benefit of the anthropologist in her presence. Tapika's desire to be identified as a good fieldworker trying to convince a difficult research subject to participate correctly in this activity performs her absorption of the project's vision to collect accurate and precise data (Madiega et al. 2013, 23). Yet Tapika's effort to translate nyembanyemba in a standard and normed fashion intersected with the contours of her unfolding social encounter with Josiah. The promise of nyembanyemba to collect high-quality, more accurate information about rural Malawians' subjective expectations was in ongoing tension with the difficulties interviewers faced in implementing the exercise precisely, that is, in a standard and consistent manner across respondents. A culturally relevant tool, then, is encumbered by the coconstruction of culture itself. In touching, manipulating, and debating the beans—a material technology validated by demographers across many research contexts—a close reading of Tapika and Josiah's encounter exposes accurate data as inherently cooked: the numbers scrawled on the survey page

and subsequently aggregated with those supplied by other respondents to other interviewers are not stand-ins for reality but rather provisional and improvised artifacts of a social negotiation.

Tapika and Josiah's interaction with the beans recalls the well-known metaphor of bean counting or the bean counter, a phrase that refers typically to a person who is excessively concerned with accounts or figures, often to the detriment of other aspects outside the figures and numbers. The act of counting beans, tiny tokens with minimal to no value, also carries the negative connotation of misplaced focus, a metonym perhaps for global health experts' uncritical investment in numbers as the sole or most important measure of efficacy and success (Adams 2016a; Erikson 2016). Tapika takes up LSAM's mandate to count the beans, but the frictions that arise between her and Josiah during the research encounter reveal the absurdity of the activity and foreground how bean counting, rather than accessing true probabilities held in Josiah's head as it seeks to, is reduced to child's play in his and other participants' eyes. Yet, Tapika—the reluctant bean counter in this scenario—makes every effort to ensure each bean is counted for the sake of the quality of LSAM's data.

Bean counting has not always carried its familiar negative connotation. Bean ballots were common to colonial New England elections, for example, when people voted with "Indian beanes" or black and white peas for their desired candidate (a practice likely imported from England), and bean counters were people of demonstrated integrity ("The General Laws and Liberties" 1672; Bishop 1893; Gross 1898; Leonard 1954). In ancient Greece, "pebbles" of "small, thumb and finger size" were the quintessential symbol of Athenian democracy. Beans were used whenever there was recourse to counted votes and in law courts when voting for the plaintiff or defendant (Netz 2002, 337; Everson 1996). In one Athenian practice, the beans themselves elected candidates via a randomization device called a *kleroterion* that had two columns with individual vertically stacked slots. Plaques with candidates' names were arranged and dropped into the slots on one side. Into the other column were dropped balls, some black and some white. As the counting machine dropped out a name and a ball in parallel, white would indicate the person had been chosen, and black would disqualify him from election. The "beans," then, acted independently of human agency to control the results of important elections, minimizing the possibility of human corruption tainting a democratic system (Netz 2002, 337). The nyembanyemba exercise, even as it aspires to collect the cleanest and most accurate data related to a respondent's felt probabilities (thus positioning the fieldworker as bean counter), is not unlike the klero-

terion; in practice, the beans exercise often seemed to resemble a divination session, throwing the bones, or casting lots more than a scientifically validated tool for collecting better data.

PROBING FOR THE TRUTH

Another important dimension to collecting accurate data entails ferreting out lies or false information provided by respondents, and ensuring that no blank space is left on a survey page. The main technique employed to achieve these objectives is probing, or *maprobing* as interviewers and supervisors termed it. Probing, or fishing for more information than a respondent initially provides in response to a question, is a key skill for good interviewers to cultivate. During project training sessions, fieldworkers were taught how to avoid being cheated by respondents who might lie or feign nonknowledge for various reasons.[6] Richard, a supervisor, cautioned the interviewers on his team: "Watch out for contradictions, or things that seem illogical, like, 'I'm twenty-one and I have six children.'" Such warnings encouraged interviewers to be vigilant seekers of the truth.

In particular, LSAM emphasized the importance of probing to ensure accurate responses to sections of the survey focused on assessing the economic shocks (Section 5) experienced by a household in the past five years (death, illness, poor crop yields, loss of income, natural disaster), listing individuals a household might seek help from in the event of future shocks (Section 6, "Potential Transfers Roster"), and listing the actual individuals a household received assistance from in the past two years (Section 7, "Actual Transfers Roster"). Patrick, the American LSAM fieldwork manager, told interviewers, "We want to see economic shocks [recorded on your surveys in Section 5] because research shows they happen. Don't leave this section blank. Probe!" Later, in reference to Section 7, where respondents were asked to list the names of up to ten individuals who actually provided them with financial assistance in the past two years, Patrick again emphasized the importance of probing: "If they say they haven't gotten help from anyone in two years, you know they are lying. You are Malawian." Similarly, he discouraged interviewers from using Code 24 ("did nothing") in response to the question, "Munachita chiyani pofuna kuthana ndi vutoli?" (What did you do to overcome this shock [problem]?): "You *know* 'doing nothing' is not what happened!" A good interviewer, he suggested, should use this code very sparingly and only after serious probing failed to uncover the answer. "It's better to have something there than nothing," he said. In the advice presented here, we note that uncovering lies presumes a

kind of local expertise and local origins ("You are Malawian"), consistent with the constructions of local knowledge elaborated in chapter 2. The supervisors, in this case, traffic in advice that furthers the epistemic investments of projects themselves: collecting accurate data by ensuring a completed survey contains no blank space or false information (see West [2016, 92], who documents the same among health surveillance assistants in rural Malawi).

Probing was also framed as an effective mnemonic device to improve recall of information, particularly that related to age of respondent or relatives. Mba (2014, 14) notes that age falsification by the respondent, ignorance of age, or cooking of age data by enumerators have long been major contributors to poor data quality on age across censuses and demographic surveys in sub-Saharan Africa.[7] Interviewers were taught how to deal with respondents who claim they are unable to recall their (or their child's) year or month of birth. To zero in on a date to fill into the survey, an interviewer could pose a variety of probing questions, such as whether they were born around independence (1964), whether their child was born during a harvest month, whether it was cold outside (June–August), and so on. However, even amid such probing efforts, evaluations of age-related data drawn from multiple African national censuses indicate that both male and female respondents preferentially report ages ending in zero or five, throwing into question the truth captured by such techniques (Mba 2014, 23).

In training sessions, probing was cast as a key technology for collecting true information. Before commencing real interviews, trainees were expected to try their hand at survey administration during pilot surveys, which had the dual purpose of piloting the surveys to catch mistakes in the content or linguistic translation, and of piloting the interviewers themselves to determine whether they were able to individually reproduce the collective standards for data held by the project. This liminal period between the completion of training sessions and the commencement of full field research was a time of significant anxiety—a rite of passage—for potential fieldworkers, who understood themselves to be under close scrutiny by supervisors and project leaders. Fieldworkers aimed to masterfully perform the skills and techniques they were taught a few days before, and to return with a neat, complete survey in a reasonable time frame. During the liminal period before potential fieldworkers transitioned into full-fledged employees, they sought to showcase their competence in the interest of earning a job for the next few weeks or months.

On one occasion, an LSAM fieldworker sought to pilot the survey at a household where the respondent refused to answer two sections of ques-

tions, and was forced to return to the minibus with an incomplete survey. He insisted to the supervisor that he probed tenaciously but was thwarted by the respondent's staunch refusal to answer his questions. So worried was he about losing his job that he pleaded with his supervisor to accompany him back to the household so he could prove that the blank space on the survey was the fault of the respondent, and not a symptom of his poor interviewing skills. Interviewers' economic interest in earning a daily wage for the duration of a fieldwork season motivated them to internalize and attempt to embody the expectations and standards for data collection to the best of their ability.

Importantly, the rhetoric of probing and lying respondents positions the research project as endowed with the ability to see or make visible the truth, betraying a primary investment in collecting data that are representative of an imagined authentic rural social reality, a reality that is always already prefigured by the questions that capture it. Research encounters were also imagined as mimicking or reproducing a real-life conversation; supervisors often emphasized to their interviewers that probing is a way to show a respondent "you are really listening, and not just recording information down on a paper." But seeing like a research project circumscribed the nature of this chat. Interviewers soon discovered that some responses provided by respondents did not easily fit into the options, codes, or boxes provided by the tangible survey in front of them. After piloting, interviewers for MAYP pointed out that some of the survey questions did not allow for commonly given responses. For example, one question asked which district in Malawi the respondent and family had originated from. Since a common response was "Zambia"—a neighboring country—fieldworkers complained about the built-in limitations of the survey (MAYP later added a note: "Record country if not born in Malawi").

On a household roster for MAYP, interviewers were asked to insert the appropriate code next to each listed name to indicate relationship to the respondent. Interviewers argued that the code "1: Husband/wife" did not sufficiently capture the relational category "cowife." Though they coded this response as either "1" or "12" ("other relative"), they suggested that it surfaced so frequently as a response that it deserved its own code.[8] Similarly, LSAM interviewers suggested to their supervisors that a code be added for the third of a four-part question about the number of sexual partners the respondent's best female married friend had in the past year (see figure 4.3). When asked a question about how they knew how many sexual partners their friend had, many respondents responded idiomatically, something like, "She was caught red-handed." When fieldworkers suggested this be added as a code for the question, LSAM researchers generally agreed that the questions should be

S19c Mukudziwa bwanji kuti anagonanapo ndi amuna amenewa? *How do you know she had sex with these partners?*	She told me.....1 Saw her coming & going.....2 Rumours/other people told me.....3

FIGURE 4.3. Question S19c from the LSAM questionnaire, 2008.

amended to improve accuracy. However, as Patrick, the American demographer managing fieldwork at the time, explained to the supervisors who brought him these suggestions, "We can't add a code without messing up things in terms of the past, data we have already collected. We must keep the phrasing and translation of the questions consistent, even if they aren't the most accurate. It's too late. . . . In order to measure change, we have to ask things in the same exact way. We have to have the same codes every wave even if they're not correct. So, just fit those responses [i.e., those mentioned above] into the existing categories."

Seeing like a research project, in this regard, necessarily implies a certain conservatism of vision. Patrick's suggestion that altering codes or phraseology of questions in the present would "mess up things in . . . the past" indexes a tension between accuracy (collecting the most true answers) and reliability (collecting such answers in the same way year after year). In his words, we note that, rather than collecting the most complete picture of rural social realities, research projects collect data that are always already, and self-consciously so, incomplete and incorporative of errors. Interviewers' embodied decisions and negotiations in the field reconcile the gap between sometimes dueling epistemic investments (accuracy and reliability) and place the onus of clean data on interviewers. The probing skills so valued by researchers are key to collecting the truest data, but, in Malinowski's ([1922] 1984, 192) classic words, these data may not be "full-flavoured" but "squeezed out of reluctant informants as a trickle of talk." In reflecting on the effects of probing on data collected by fieldworkers on projects in five sub-Saharan African countries, Randall et al. (2013, 780) echo Malinowski: the interviewer "extract[s] data from the respondent" and may "make respondents say things they had not thought about or possibly do not want to say." As is evident in the case of LSAM's questions about household shocks and probabilities, respondents fulfill their role in research by simply providing a plausible answer, as arbitrated by the fieldworker they encounter.

	SUPERVISOR	LOGGED BY	CHECKED BY	ENTERED BY
INITIALS	_____	_____	_____	_____
DATE	_____	_____	_____	_____

FIGURE 4.4. The life course of data (taken from LSAM questionnaire first page, 2008).

PERSONALIZING DATA: CALLBACKS AND CHECKING

The plausibility of respondents' answers, however, is meticulously measured and technically mediated at many points along data's life course. Figure 4.4, taken from the front page of the LSAM survey, shows the path that data take as they are manufactured, and whose hands they pass through as they are converted from raw information (survey responses) to valuable data (statistics derived from a database of good numbers). When an interviewer returns to the field supervisor with a completed survey, the supervisor checks the questionnaire to ensure there are no immediately obvious inconsistencies between responses and no missing information or blank spaces. If the supervisor discovers missing information, the interviewer is sent back to the household to collect it. Once the survey questionnaire is deemed complete, the supervisor initials and dates it. It is carried back to the field office in the minibus and deposited in a "to be logged" box, where it waits to be logged by a data entry team member. Next, it is checked again by a data clerk; if inconsistencies are discovered, it is sent back to the field the next day to be corrected by the initial interviewer via a callback. Finally, when a survey is deemed complete, consistent, and credible, it is entered into the database by a data clerk. At this point, the survey has passed through many hands, indicated by the differently colored pen marks and initials on the front page. After being logged and entered, surveys are scanned and archived in boxes labeled with village numbers.

The product of all this labor—clean data—is valuable precisely because it passes through so many hands. The initials scrawled on the front page not only signal the phases through which data pass but also point to the logic and mechanisms of seeing like a research project. No one person arbitrates the quality of data; instead, a number of individuals whose habits and ways of seeing have been harmonized (to various degrees) all claim ownership over data at one point in time. In the snapshot of data's life course in figure 4.4, we see how the different cogs that constitute the machinery of the research project are supposed to work together. Even if data in their final form—statistics or numbers derived from a database—may appear to be abstract and unanchored from their origins at a rural household, data are personalized at every

stage, even in their raw forms. Notably, however, the persons who handle data along their life course are typically "the most poorly paid and least well qualified link[s]" in the data production process even as they, in reality, might have the most influence on the value of data collected (Randall et al. 2013, 784).

This personalization works as a push and pull mechanism to ensure standardization and collection of high-quality data through self-surveilling checks and balances. In the process of fieldwork, the callback acts to clean data as they pass between an interviewer and supervisor. Supervisors who discovered inconsistencies in the pages of a recently collected survey would summon the interviewer to question the inconsistency. If there was no viable explanation, the interviewer was sent back out to the field to revisit the same respondent to find out the truth. Callbacks were loathed by interviewers, who put in extra effort during a first research encounter with a respondent to avoid having to revisit a household. Having too many callbacks marked an interviewer as incompetent or lazy and put one's job in jeopardy.[9]

Esau, an LSAM supervisor, discovered an inconsistency in the way an interviewer called Edward on his team had recorded information about the number of children his respondent had in his first and second marriages. Esau suggested Edward find the respondent and ask the question again to be certain he got the numbers correct. Edward protested and, since it was dusk, it was decided he would go collect the information the next morning. A few hours after fieldwork began the next morning, Edward returned with his survey; he had neatly crossed out the number given for children in a second marriage and replaced it with the correct response. However, upon receiving the callback survey, Esau accused Edward of cooking the number, implying he had made it up. On our way out of the village at dusk that day, Esau ordered the minivan to stop near the trading center where the respondent in question was based. He inquired directly with the respondent whether Edward had revisited him that morning, and the respondent indicated he had not. Edward sat in the back of the van, shamefaced and quiet as Esau chastised him in front of his team members. As we disembarked at the field office half an hour later, he exclaimed, out of earshot of Esau, "*Eeee* . . . they [supervisors] don't know how difficult it is to make someone sit for hours asking them questions, and then to go back again yelling 'Hodi!' [standard Chichewa greeting used to request entrance or announce arrival at someone's compound or home] a second time. . . . You become a laughingstock."

The negotiated friction between Edward and Esau indicates that data's travels are circuitous and do not follow a straight path. Lying—most obviously framed as a common practice among rural respondents—manifests

also as a deviation from standardized habits and practices entrained into fieldworkers as Maussian techniques of the body (Mauss 1973). Even as lying appears in the final data as gaps, error, or messiness, it operates tactically in the field as it does in everyday social life (Salamone 1977; Bleek 1987).[10] Peterson (2002, 388), for example, in an examination of colonial census data collected in French postwar southern Mali, suggests that respondents were motivated to declare the religion they thought was the safest vis-à-vis outsiders and the state, which led to a false significant increase in Muslims counted. Recall, as well, how in chapter 3 we saw how people impersonate individuals in the sample to receive the soap gift projects distributed to their respondents.

LEARNING TO WRITE (AGAIN): HARMONIZING INTERVIEWERS AND MUTING INTERVIEWER EFFECTS

> Working in the field is determined by the field itself. . . . You can't plan from the office the things that will come up out there. —Chifuniro, LSAM supervisor

Chifuniro's advice to a new crop of interviewers points again to the discursive spatiotemporal boundary drawn between the clean and orderly office and the messy and unpredictable field. While one can plan for and attempt to predict the impediments and challenges to be faced during data collection, it is nonetheless an endeavor determined by the field itself. In this way, prefieldwork training sessions for interviewers seek to, as far as possible, mitigate what demographers term interviewer effects, or measurement error due to interviewers' characteristics or practices in the field. Training sessions are crucial moments of standard setting, where the project seeks to establish uniformity in fieldworkers' comportments and practices by introducing a set of agreed-upon rules for data collection (Bowker and Star 1999; Timmermans and Epstein 2010).

Of course, it is difficult to control for effects that an interviewer's age, gender, or ethnicity might have on a respondent's answers in a given encounter.[11] Nonetheless, projects employed tools to attempt to document and measure such effects. The last page of LSAM's survey was an interviewer's questionnaire, which directed the interviewer to answer eight questions "soon after the interview." These questions were meant to capture (1) potential role-independent interviewer effects (e.g., social identity) on the course of the research encounter or the data it produces; and (2) potential measurement error due to lying respondents. First, interviewers were asked to rank the respondent's physical attractiveness relative to other persons of about the same age and sex on a scale of 1 (much more attractive than average) to 5 (much

less attractive than average). These questions were attempts to access the effects of social characteristics on respondents' answers. These data were then collated with data collected from project supervisors, who were asked during fieldwork trainings to rate the attractiveness of all the interviewers on the same 1–5 scale.

The interviewer's questionnaire also served as a system of checks and balances on prior responses to the same questions asked during the course of the survey. For example, interviewers answered the question: "Does the respondent's house have a pit latrine?" (yes, 1; no, 0). In bold capital letters, a note to the interviewers compelled them to check for themselves to determine the answer (in trainings, interviewers were encouraged to request to use the toilet at some point during the interview to subtly ascertain whether or not a latrine was present).[12] This (like trap or red herring questions embedded in questionnaires, as well) was meant to check whether the respondent had answered the same question posed back on page 2 accurately, or had lied, which might call the truth of other responses into question also, a built-in means of determining potential measurement error at the level of a respondent (Kasprzyk 2005, 172–173). The interviewer questionnaire collected a set of mimetic metadata that implicitly indexed patterns of response variation and added value to the data collected (Vemuri 1994). The auxiliary data collected in a survey that help describe the data collection process are commonly referred to as "paradata" in demographic parlance, and reveal researchers' massive investment in monitoring data quality.

In training sessions, research projects focus on minimizing measurement error due to role-restricted interviewer effects or differential response patterns that might result from interviewers' different interviewing styles, their differential adherence to guidelines or survey scripts, and so on (Sudman and Bradburn 1974; Stecklov and Weinreb 2010). Relational politics between interviewer and interviewee, too, can affect data collected: Loveman (2007, 91–96) analyzes data from the 1910 and 1920 rounds of the Puerto Rican census to show how interviewers brought assumptions to bear in their classificatory decisions around the race of their respondents, with a "whitening" effect on the census data. Interviewers' primary labor during fieldwork is recording responses with pencil on the pages of questionnaire after questionnaire, long the dominant mode of data collection in developing countries and known as paper-and-pencil personal interview. In recent years, survey projects have begun implementing tablet- or smartphone-assisted personal interviewing, but at the time of this research, interviewers used paper questionnaires. Before pencil goes to paper, however, interviewers relearn how to write. In addition to

writing neatly, interviewers must ensure they leave no blank spaces, follow the script and instructions of the survey meticulously, and accurately translate raw information into the appropriate codes.

The training sessions discussed in chapter 2 were the primary site in which projects sought to harmonize the behaviors and writing practices of project interviewers. Clean surveys necessitate that all interviewers record information (or lack of it) in the same fashion. Learning to write (again) was a long process that entailed going through the survey questions one by one and painstakingly providing specific instructions on how to properly record information. On the first page of the survey, where interviewers were meant to record accurate information about the respondent, including age, birthplace, father's name, and so on, no codes were provided; rather, the interviewer had to neatly write out given answers. As we moved through the survey pages in a training session, we often paused to reach consensus about how to record information consistently: "We should now agree that instead of leaving a blank space on the first page, we must write in a dash instead," with the trainer drawing a dash on the flip chart at the front of the room.[13] The directive was followed by a question that functioned to index consensus and progress throughout the training sessions: "Eti? Onse pamodzi?" (Is it so? Are we all together here?) The fieldworkers' ritualized response ("Eeee! [Yes!]") cements the solidarity and collectivist orientation—centered on the research project's-eye view that is the imperative of these trainings (Vertesi 2012, 405).

Training sessions also familiarized interviewers with skip patterns in the survey. Skip patterns—highlighted by instructions for the interviewer embedded in the survey itself—chart a course for interviewers as they proceed from question to question and page to page of the survey in the field. Two skip patterns, indicated by >> (see figure 4.5), direct the interviewer to proceed, in the first case from question 10 to question 12 and, in the second case, from question 11 to question 13. In addition to observing the skip, interviewers were expected to treat the blank space produced by a skip in consistent and harmonized manner. While some projects encouraged interviewers to mark a skipped question with a dash, others taught interviewers to leave a blank space.

For the numerous interviewers and supervisors who worked on many different research projects, these idiosyncratic preferences were difficult to master, and their relevance to the quality of data collected often opaque. Supervisors often absolved themselves of responsibility for these guiding rules, using the term "azungu" (Chichewa term for foreigner) to emphasize the fact that such directives came from above and were out of their hands: "The

10.	11.	12.	13.
Do you have electricity working in your dwelling? 1= Yes 2= No (>>12)	Is your electricity from ESCOM, a generator, solar panels, or some other source? 1= ESCOM 2= Generator 3= Solar panel 4= Other, specify (>>13)	Although you do not have electricity here, is there electricity within 100 meters of this dwelling, whether from ESCOM, a generator, a solar panel, or some other source? 1= Yes 2= No	Is there a landline telephone in working condition in the dwelling? 1= Yes 2= No

FIGURE 4.5. A skip pattern embedded in the LSAM questionnaire.

azungu do not want you to write any leading zeros, so do not do it" or "The azungu want us to use the code 'Other' as sparingly as possible, so avoid it." Despite the emphasis on observing skips and the ample attention paid to writing practices in the trainings, they became a source of much frustration for interviewers during survey administration in the field. During fieldwork pilots, many interviewers felt discouraged by all the red ink on the pages of their completed surveys, indicating their supervisors' many corrections. The red marks identified errors of content (e.g., inconsistencies between responses or responses that were not likely true) or errors of form (e.g., where interviewers had failed to observe skip patterns, used the wrong marks to signal a missing response, neglected to ask a question, or written sloppily).

Often, techniques to ensure accuracy and techniques to ensure harmonization and efficiency came into friction. As discussed at length above, probing is a valuable skill for interviewers to cultivate and helps ensure that interviews flow more naturally and take the shape of real conversation rather than sterile survey encounter. During training sessions, interviewers were encouraged to think of the questionnaire as a form of chatting (*kucheza*); however, this often threatened the quality of data collected by increasing the influence of interviewer effects on data. In early June 2008, Ishmael, an LSAM interviewer, explained that he most enjoyed administering the vignettes section of the questionnaire. "I have fun with them because I like to tell the story in my own way," he suggested. The vignettes, inserted into the survey by a sociology graduate student, were meant to measure a respondent's perception of agency as played out in fictional stories constructed to have local relevance and solicit data of value to researchers (see figure 4.6).

V5	Rose is married to a man who moved around with [had sexual relations with] a girlfriend for many years while they were married. When she found out, she told him to stop seeing the girlfriend or she would divorce him. He stopped. How easy is it for Rose to protect herself from getting infected with HIV/AIDS?	Very easy Easy Difficult Very difficult Don't know	1 2 3 4 88
V6	Beatrice caught her husband "red-handed" having sex with another woman. She took her case to the ankhoswe [traditional marriage counselors/advisers] and said she wanted a divorce because she was afraid of getting HIV/AIDS and she was no longer able to trust her husband. The chief granted her the divorce and she didn't have to pay any money. Beatrice went through with the divorce despite her husband's protests, and she returned to her parent's home. How easy was it for Beatrice to protect herself from getting infected with kachilombo [HIV]?	Very easy Easy Difficult Very difficult Don't know	1 2 3 4 88

FIGURE 4.6. Two vignettes from LSAM questionnaire.

Though Ishmael was widely known to be a well-performing interviewer, his supervisor reminded him that it was essential he read the vignettes exactly as they appear on the survey page, to ensure that all respondents hear the vignettes in the same way (and thus to mitigate measurement error resulting from role-restricted interviewer effects). This advice betrays the project's interest in collecting timely data; interviewers were left to negotiate a small space between administering a questionnaire like a chat and collecting complete data as quickly as possible.

On one visit, I accompanied Janet, a twenty-six-year-old female interviewer, to her meeting with a thirty-nine-year-old woman called Namoyo. When we arrived, Namoyo and her mother were shelling maize on the *khonde* (verandah). Before getting down to business, the four of us sat quietly together, each working at the maize. Maintaining our place on the khonde and continuing to

shell maize, we began the questionnaire. Now and then, children, goats, and chickens darted across a walking path nearby, disrupting the flow of the survey. Janet introduced the survey as an informal chat: "Naphiri [my Chewa name] and I are just here to have a chat with you!" In both English and Chichewa, kucheza (to chat) implies conversing in an informal, nonlinear, undirected, and non–temporally bounded manner—free-forming a conversation. But as soon as Janet brought out the questionnaire and her pen, it became evident that this particular chat would closely follow the order of the questions written on the survey pages.

The first portion of the chat involved Janet verbally eliciting and carefully filling in the household roster (for a sample roster, see appendix). This roster was a table with fifteen columns and ten rows. After asking Namoyo to list each member of her household, Janet wrote the names one by one into the blank rows. Once all the names were recorded on the sheet, she asked a series of questions about each household member: "How old is X? What is X's relationship to you? Is X's mother alive? In what year did X move here? What is the highest level of schooling X went to? Is X married? Is X ill?" Many of the answers provided by Namoyo had to be coded by Janet with a relevant number. In cases where she did not recall the codes, Janet paused the chat while she leafed through an accessory packet of questionnaire codes in order to find the proper one. A month earlier, Janet had attended a training in which project interviewers had been taught to maintain good penmanship and be careful and consistent in filling out project surveys. As Namoyo delivered her responses to the survey questions, Janet took care to record the responses neatly; she even used a ruler as a straight line beneath the letters she wrote. The chat was marked by long periods of silence as Janet monitored her own penmanship to ensure she was seen as a good interviewer, not only by me but by the researchers and data entry clerks who would see the marked-up questionnaire later in its life course.

Despite the recipe provided by the survey from beginning to end, survey chats were certainly not linear. The encounter between Namoyo and Janet confounds survey researchers' claim that modules or sections of the survey should match the order in which the interview is to be conducted so as to mirror natural ordering (Glewwe 2005a, 41; Dillman 2008). Namoyo could not recall the names of her parents-in-law when initially asked by Janet; later in the survey, however, she suddenly remembered them, interrupting the flow of the interview session and prompting Janet to flip back a few pages to enter the information. Like the rhythmic shelling of maize, the survey's chronology served as a mere backdrop against which our interaction meandered. The

interview encounter was a negotiated space of flows and stoppages of data symptomatic of the interests of the interviewer and interviewee, respectively. As was the case in most of the interviews I observed, the interview between Janet and Namoyo was marked by interlocutors' mutual testing of the waters.

Early on, Namoyo commonly responded to questions with "I don't know," or by providing other noncommittal answers. When Janet asked her about the amount of money she loaned to others in the past year, she claimed "none." Janet looked at her dubiously, laughed, and probed, "Not even five kwacha [about 4 cents USD at the time]?" Namoyo laughed, and then agreed that she had indeed loaned friends, neighbors, and family members money in the past year. Later, Janet had to return to this box on the survey again when it turned out that Namoyo could remember the amounts she donated to individuals she listed by name. Similarly, she claimed she could not remember the ages of her own children. When Janet pressed her, she could.

Finally, over the course of a series of questions that covered wealth indices, Namoyo grew frustrated and visibly annoyed at having to provide verbal responses to questions that she felt were self-evident to Janet. As a good interviewer who had been taught never to miss a question, Janet enunciated each question: Does your household own a TV? Solar panels? Does your household have a metal roof? Namoyo laughed in the face of such questions: Janet could easily see that she possessed none of these items—she was poor! Yet when Namoyo laughed, Janet still pressed her to verbalize her actual response: "No."[14] Often, respondents' ambivalence about participating in a survey aligned with the interviewers' own ambivalence about the agenda of the project that employed them (May 2008). Janet's affective orientation to Namoyo's sighs of frustration showed that these questions were not her own; she made it clear that she was merely a mouthpiece for LSAM. Namoyo, picking up on Janet's apparent disinterest, made repeated stabs at taking control of the interview encounter by being selective about which questions she answered, by providing inconclusive or vague responses, or by feigning nonknowledge before finding an answer. These efforts tested the contours of the interview as a social space: How invested was Janet in securing answers to each of the questions? How much could Namoyo reveal? Was Janet able to detect when Namoyo provided bad information?

In the space of the formal survey, Namoyo relished the chance to talk to Janet and me; as outsiders, we were a valuable and novel source of information. Namoyo asked us how things were in other districts to which we had traveled with LSAM, whether we had any children, and so on. Again, the linear form of the survey meandered when it was inserted into the social

relations and space of the interview encounter. The standards and guidelines for collection of good numbers that interviewers learned in training sessions translated into the field in unpredictable ways through the instrument of the fieldworker (Mauss 1973). The imperative to write neatly appeared in the field as awkward silences, with goats bleating in the background and informal conversation filling the gaps. The mandate to ask every question became the site of a negotiation, with both interviewer and interviewee trying to gain a foothold to express and secure her interests. The command to leave no blanks on the survey prompted push-and-pull exchanges between Janet and Namoyo, with the former probing for pieces of information and the latter recalcitrant about providing it. The chronological time presumed by the numbered pages of a survey and the project's emphasis on efficiency and timeliness were enacted by Janet's careful administration of the survey but came into friction with both her desire to be a good interviewer (which often involved slowing down to record data well) and her circuitous and slow time encounter with Namoyo.

OPTIC TECHNOLOGIES: POLICING AND PATROLLING THE SAMPLE

Producing high-quality data presupposes meticulous sampling strategies. It is impossible for projects to interview all Malawians, but in order to attain high-quality numbers, a sample must include a large enough number of households to support the eventual claims made from the data. Beyond ensuring that the absolute number of sampled households is sufficient to ensure that data will be of high quality according to epistemic investments in statistical power, projects must also protect sample purity; the sample must capture not only ten individuals living in sample households, but the correct ten individuals.[15] In demographic terms, this entails interviewing the same individual year after year. Panel survey projects must minimize threats to sample representativeness that may arise from how a sample is chosen and followed over time, participation rates in a survey, and the procedures of data collection.

Like researchers, fieldworkers were well aware of the importance of both sample size and sample purity, which manifested in their everyday embodied practices as concerted efforts to locate and successfully interview all individuals in the sample. Over time, they came to see the sample as a cohesive whole, even as they interacted with its individual members on a daily basis. For example, in discussing the importance of properly introducing research project objectives to local traditional authorities such as chiefs, supervisors told their fieldworkers, "We must respect the sample at all times." Seeing like a

state (or a research project) entails deploying a set of techniques and tools designed to guide people's conduct as individual units of a population (sample). Much like Foucault's ([1978] 2007, 137) metaphorical shepherd cares for his flock, research teams, too, care for the research sample from birth to death over longitudinal time. The sample is an organizational and, as will become clear, a political unit. Whereas a top-down view of the sample might suggest that its individual members are interchangeable, a bottom-up view indicates quite the opposite: producing high-quality data relies on the systematic collection of freely given information from thousands of individuals, each enmeshed in complicated social networks, each with a unique geographic location, and each with his or her own agenda.

Recalling James Scott's (1998) elaboration of the efforts of eighteenth- and nineteenth-century German scientific forestry to manage and order the ecological messiness of forests, we note the important role that tools and technologies such as maps, devices to measure tree size, and surveys played in allowing the state to narrow its focus or vision to see only what it wanted to see: the revenue from timber extracted annually. Survey research in the age of global health employs techniques not unlike those taken up by the state in census or other national projects. Technologies of enumeration make visible slices of reality that are of interest or valuable to a particular situated gaze.

Survey researchers are well aware of the detrimental effects of attrition—failure to find or reinterview individuals who were surveyed in earlier waves or visits to the field—on the quality of their data. Attrition leads to a decreased sample size that reduces power in statistical analyses and is a major factor in poor data quality in sub-Saharan Africa (Alderman et al. 2001; Bignami-Van Assche et al. 2003). Mobility of respondents, failure to find respondents, and respondent refusals are major threats to data quality. Epistemic investments in sample purity and sample size translate on the ground into various techniques and tools that help a project efficiently and effectively see a sample. Emulating the unfolding relations of the field, I provide two ethnographic vignettes to bring to life how the sample is bounded and how data's purity is maintained by the improvised and unscripted practices of fieldworkers.

Even as research teams come to see the sample as a single entity, it is a living, breathing organism whose shape-shifting nature perpetually threatens to exceed or escape the gaze of the research project. To combat sample attrition, field teams are outfitted with an arsenal of instruments meant to allow them to see and keep the sample pure on a daily basis. The movements, meanderings, and interests of respondents, however, challenge these optic tools.

Nonetheless, these instruments allow the project to patrol the borders of the sample sufficiently to produce numbers that are good enough, if not perfect.

> The large table in the common room of the MAYP field office is littered with hundreds of envelopes. These are the tool kits of the fieldworkers who will collect data tomorrow from villages within fifty kilometers of the office. Each contains a set of tools necessary to locate and interview a single respondent; indeed, the contents of each packet stand in for and create a sort of enumerable person. Within each folder is a color photograph of an adolescent male or female. I remove one and see a boy, age seventeen, in a blue T-shirt standing against a backdrop of bricks likely to be his home. He squints into the sun and holds a white paper across his chest that reads 102_Madumbo_34. This placard indicates for the interviewer assigned this boy the enumeration area, traditional authority, and household number (e.g., EA_TA_HH). A small white sticker below the photo lists the name, sex, schooling status (in or out of school), and nickname of this respondent. Aside from the survey itself, the envelope also contains a map, hand drawn by the fieldworker who visited the same respondent last year. The maps include landmarks ranging from trees to shops to football pitches [soccer fields] and churches, represented in the unique hand of interviewers. These maps—which capture reality from the perspective of an observer on the ground—contrast with the large official maps kept in the field office and are complemented by the GPS coordinates of the household, included on the same page as the map. The teams were provided with bright-yellow, heavy-duty GPS devices, though they were rarely used in practice.[16]

This excerpt from my field notes highlights the arsenal of optic tools utilized by projects to bound and see realities of interest: photographs, enumerative labels, GPS devices, and hand-drawn maps. In fact, each envelope stands in as a proxy for a real person, valuable to the project as a coherent unit of data. Here, we see how much the project already knows about its subjects, and come to understand the labor and technologies invested in finding and successfully surveying each of these individuals.

The technologies for locating respondents are numerous, but nonetheless respondents employed tactics, either deliberately or by virtue of being absent, to escape or evade the project. A respondent could be out: working in the *dimba* [wetland garden] or trading center, on a trip to South Africa, in the city, at the hospital, completely relocated to another residence outside the

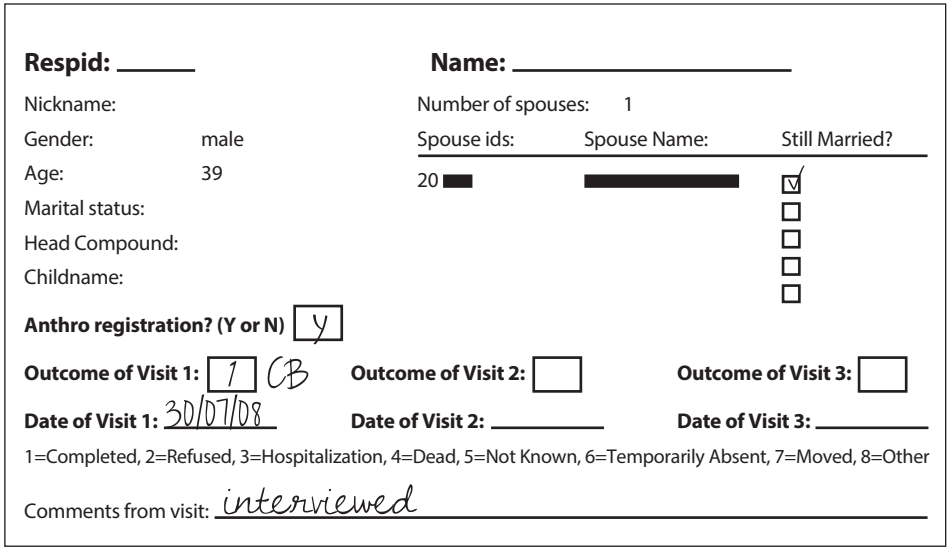

FIGURE 4.7. Log form for recording interview outcome, LSAM.

spatial bounds of the sample, or, in some cases, passed away since the previous interview. When an interviewer arrives at a household and hears "Alibe! [He/she is not here]," the interviewer proceeds to record the reason on a log form with eight possibilities, depicted in figure 4.7.

Each outcome entails a series of next steps that illustrate how finding respondents—even those who have disappeared—is a cat-and-mouse game. For example, if it was determined that a respondent was deceased, the interviewer proceeded to administer a mortality questionnaire, also called a "verbal autopsy" by the research project teams.[17] The verbal autopsy entailed interviewing a family member or other person close to the deceased to, as closely as possible, ascertain and document the cause of death of the respondent. Even in death, then, respondents did not escape the gaze of the project; their movement out of the sample needed to be documented to preserve the integrity of the sample. Each respondent designates, from the view of demographers, interchangeable lives and deaths that somehow belong to the research project (Stevenson 2014, 27). Death may be beyond the reach of biopolitical power, but it is not outside the view of statistics: "Power has no control over death, but it can control mortality" (Foucault 2003, 248).

In many cases, a respondent was temporarily away, and an appointment could be set for the following day for a return visit by the interviewer. If

Characteristics of the Core Respondent

a. Has the core respondent moved permanently or temporarily? [__] 1= Permanently
2= Temporarily
b. Where does [...] currently live?
c. District (or country): _____ TA: _____
d. Village/town: _____
e. Area type [__]
1= Major urban 2= Boma 3= Rural
f. Head of compound: _____
g. Name of household head: _____
h. What is the nearest market or trading center? _____
i. Is there a landmark close by to where [...] stays? (such as a school, junction, etc.)
_____
j. Approximately how far is this location from here? (indicate the main means of transport as well as the approximate time, and/or distance)_____
k. What is {...}'s marital status? [__]
1= Single →Q3e 2= Married
l. Name of spouse:_____
m. Why did [...] move?
1= To work or look for work 5= Following new spouse
2=To look for land 6= Don't know
3= School 7= Other, specify_____
4= Following parents
n. What was the approximate date of [...]'s move out? (note: should be after ~June/July 2007)
Month: [___|___]
Year: [___|___|___|___]

FIGURE 4.8. Tracking form for absent respondents, MAYP.

interviewers were told a person was in the fields or at the trading center, they would walk or take a bicycle taxi to search for the respondent, and perhaps interview him while he was ironworking, farming, or selling mobile phone airtime units. Some respondents were away more permanently due, for example, to relocation or migrant labor. During tobacco season, many men sought casual labor planting and harvesting, and would live away from home for a number of weeks or months; a person might also be away closer to home engaged in *ganyu* labor.[18] In figure 4.7, we note that this particular respondent was "interviewed" (a successful outcome). In many other cases, however, respondents had relocated to South Africa since the previous survey wave ("6=temporarily absent"). Indeed, migration is the main reason for sample loss or attrition in Malawi (Anglewicz et al. 2009). In such cases, the project would complete a tracking form (see a portion of MAYP's version of a track-

ing form in figure 4.8), collecting as much information as possible about the whereabouts of the individual: who he is staying with and which neighborhood he is living in, for example.

At the close of fieldwork, teams would use these forms to track respondents, using piecemeal information collected from relatives and friends to find them. For example, at the close of MAYP fieldwork, seventy-five respondents needed tracking, among these, twenty-two in the study district of Salima, nineteen in Lilongwe District, and three in Dowa District, covering over 11,000 square kilometers.[19] "We will find them—don't worry," Hastings, a MAYP supervisor, exclaimed when fieldworkers questioned the utility of collecting this information, showing the extent to which the project was willing to go to preserve the sample. Finally, although interviewers were encouraged to work hard to avoid refusals, some participants, as we saw in chapter 3—though in LSAM's case, remarkably few—did refuse to participate (Kranzer et al. 2008; Reniers and Eaton 2009; Obare 2010). This, too, had to be documented on the form, and interviewers were asked to record some notes on the reason behind this refusal. Similarly, respondents who were too ill to be interviewed—or in some cases too drunk on *kachasu* (a variety of locally distilled liquor popular in rural areas)—were coded as refusing.

The arsenal of tools meant to track respondents who were away worked to effectively reduce attrition in the sample; however, even finding respondents who were present was not easy. Namely, before beginning an interview, fieldworkers had to verify that the respondent was who he claimed to be. As supervisor Andrews explained, "These guys have been in our sample *kalekale* [since a long time ago]. We know them! But we have to make sure we get the right person." Maps hand drawn by fieldworkers in past years often worked to help interviewers find the households they were assigned (see figure 4.9 for a sample hand-drawn map). Sketches of miniaturized trees, churches, vegetable stands, paths, and soccer fields helped fieldworkers find their way through terra incognita, though, of course, trees or kiosks could change from year to year. Each crop of interviewers was instructed to correct or improve the maps as needed and often drew over, crossed out, and refined the maps to make them more accurate in the present. In this way, these maps from below became accumulative condensations of archived project knowledge, collaboratively created, transmitted from one generation of fieldworkers to another, and owned not by an individual but a project.

Teams often relied on word-of-mouth directions, and, in especially rural or difficult-to-navigate EAs, teams would often hire a scout. This person, assumed to be a reservoir of local knowledge about the social landscape and

Household ID: | | | |

HOUSEHOLD CONTACT FORM
(completed by Supervisor/Enumerator)

1. Household GPS Coordinates

S ☐☐ ° ☐☐ . ☐☐☐ '

E ☐☐ ° ☐☐ . ☐☐☐ '

2. Detailed instructions on how to find the Household (including Sketch Map):

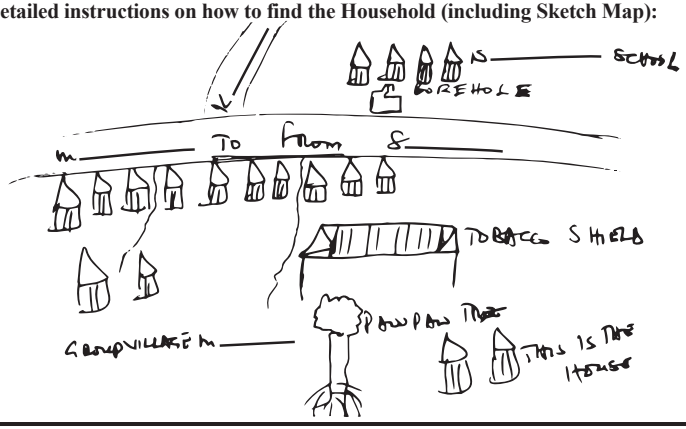

FIGURE 4.9. A map hand drawn by MAYP fieldworker (anonymous).

composition of villages in the sample, was paid a daily rate of 500 kwacha (about $3.50 USD at the time) and was often asked to locate and book appointments with respondents ahead of time to save time and ensure respondents were present when teams arrived. Scouts were often appointed by the chief of a certain area, who frequently recommended a son or other relative for the job.[20] Scouts took significant pride in their few days of employment and emphasized their status by carrying a clipboard that listed the names of respondents to be interviewed. Teams also relied on more informal channels of finding respondents, inquiring about the whereabouts of individuals by showing photos to bicycle taxis or giving women carrying buckets of water from

the borehole a ride in exchange for information about the locations of sample households. In general, the array of tools available for seeing a household—that is, making it visible against a background of village life—were very effective. On only a few occasions were respondents not trackable at all.

In coming face to face with a respondent, however, an interviewer had to verify that this individual was who he or she claimed to be. The supervisors' advice to their fieldworkers that "respondents are always trying to trick [research teams]" was sometimes borne out by interactions in the field. Names did not always work as a unique fingerprint, since relatives can share the same names or similar names. When someone claimed to be the sought-after respondent, interviewers often held up the pixelated photo next to the person's face to scrutinize the match. Often, they noted a tree or house in the background of the photo and asked the respondent, "Where is this tree?" or "Are these the bricks that appear behind you in the photo?" In some cases the numerical code of a household was scrawled in white chalk on the house itself, a visible marker that the household was in the sample. Next, the interviewers cross-checked the names, age, and nickname.

Nonetheless, a number of hiccups arose. Fieldwork teams encountered imposters, or people who would pretend to be the respondent and proceed to answer the questions. On some days, LSAM supervisors grew frustrated with the prevalence of what they called "imposter syndrome" and blamed it largely on the "hunger for kwachas" the incentives project that passed through previously had created. People posed as members of the sample because they assumed being in the sample meant receiving money or other possible benefits now or in the future. Though imposter stories became the stuff of fieldwork folklore after the fact, in the moment, imposters were a drain on time, resources, and patience. For example, Collins, a MAYP interviewer, spent one morning searching for Moses Banda, a respondent in the sample. It was well known in the sample villages that MAYP was expanding its sample that year to include spouses of respondents. When we turned up at Moses's household—according to the map in the envelope—we were greeted by Mercy, a woman who claimed to be Moses's wife. She assured us that Moses was out but would return shortly; in the meantime, in line with the sampling strategy to add spouses to the sample this year, Collins decided to interview his wife. However, about two hours later—while Mercy and Collins were still immersed in the interview—Moses arrived, and it soon became clear that Mercy was not Moses's wife, but the wife of his brother who lived in Lilongwe. Collins stopped the interview immediately, visibly frustrated at being tricked by Mercy.

The supervisors debated whether Mercy should receive the bars of soap or not, ultimately deciding to give her the gift in exchange for her time, even if the information would never become data. Mercy, motivated here by her own interest in acquiring the soap she knew would be forthcoming, pretended to be someone she was not, throwing a temporary wrench into the works of the project and threatening the integrity and purity of the sample as an interloper. As we saw in chapter 3, soap sometimes motivated respondents to pose as someone else; in a few cases a legitimate respondent did not want to answer questions and suggested a friend or relative stand in for him or her to receive the soap the respondent was entitled to. On the ground, the sample was a political and politicized unit. An optically bounded, neat and tidy entity as viewed from above morphed into a messy, shape-shifting political community on the ground, rife with spillovers and leakages (Adams and Kasanoff 2004, 344).

In the case of LSAM, the longest-running survey project I worked with, people in sample areas were very aware of who was in and who was out of the sample. Even as some people expressed frustration with the meager soap gift, there was a sense that being in the sample was better than not being in the sample, and it held a certain promise of benefits to come in the future (Prince and Otieno 2014, 940). Often, people saw the conspicuous minibuses passing through the villages and flagged us down, asking if we could ask them questions as well. The teams often promised they would see them soon, but without knowing whether these particular individuals were in the sample or not.

Certainly, the sample was the narrow lens through which both the project and its fieldworkers bounded the social reality of interest to them. In the same way that fieldworkers are taught to conceive of the field as separate from, distant from, and different from the office, the sample has to be treated in a certain way in order to ensure the collection of pieces of information in a standardized and orderly manner. Even before teams gain access to the sample, they must first engage in formal meetings with district commissioners, traditional authorities, district health officers, and local police to alert them to the teams' presence in the district for the coming weeks. The epistemic commitment to sample purity produces the sample as a thing autonomous and disconnected from the world surrounding it, an entity whose borders should be patrolled. Yet in practice, maintaining sample purity entails artfully navigating the blurred lines between "sample" and "not sample." These unpredictable and unfolding social relations between project staff and residents of

sample villages challenge the notion that research projects are alienated from the everyday realities of their research subjects. Even as the data they collect are a metaphor for the project's inability to see the "real, existing forest for the [valuable] trees" (Scott 1998, 3), fieldworkers are entangled in the social realities they aim to capture.

For example, field teams sometimes attended funerals in the villages in the project's sample to pay condolences and give monetary donations. In the event of a death, data collection might be delayed for one day while fieldworkers attended the funeral. Andrews explained to his field team, "It is our duty to show them we are part of them." Fieldworkers were discouraged from just "sitting in the minibus" and encouraged to "get to know them [people living in research areas]." This advice was largely taken up; toward the end of a fieldwork day in August 2008, a parade of women dancing vigorously to the rhythm of drums surrounded our minibus, beckoning for us to join them. The women were celebrating *nsondo*, a girls' initiation ritual practiced in Yao areas. The field team members sitting on the bus left their newspapers, conversation, and mobile phones to join the dancing. The warp and woof of rural life intersected and redirected the temporalities and prescriptions of data collection on a daily basis, and treating the sample correctly was key to collecting good data. Seemingly insignificant and happenstance encounters in the field played a key role in lubricating data collection. Fieldworkers enjoyed meals offered to them by survey respondents, engaged in business transactions with local people (e.g., purchasing honey, fruit, or local chickens from purveyors, or buying bread and tea from the same tea stand over the course of one week), gave sick people rides to the hospital, helped women pound maize, played football with young people, and so on. Each of these small interactions functioned to elongate the relationships and build trust between a project and its sample.

Knowing the trees, in this case, is a prerequisite for seeing the forest. Even as the project itself focuses myopically on the sample as the unit of value, the production of this value is contingent on forging the right kinds of relations with those within and outside that unit. Further, actions in the present can enhance or compromise the ability of the project to collect good data over longitudinal time. In many cases, this entailed ensuring proper relations of exchange and obligation were maintained. I reproduce a scene from my field notes to show how minor but tactical investments in maintaining good social relations worked to ensure data collection went smoothly (not unlike the anthropologist's own directive to build rapport).

> The MAYP Land Rover moves slowly through tall grasses, swimming in mud that lies beneath. Suddenly, we strike something hard. A man emerges from the bushes, yelling that the car has run over a clay pot filled with the day's relish (*ndiwo*). The supervisors quickly got out of the car and apologized to the man. He accepted their apology, but suggested they should compensate him for the broken pot. The supervisors consulted among themselves, and decided to give the man 600 kwacha ($4 USD at the time). The man received the money gratefully and we went on our way. Henry later explained the story to the researchers back at the office, and was given the 600 kwacha he paid the man out of his own pocket.

Here, the researcher validates the supervisors' decision to compensate the man for the pot, even though the broken pot was technically no one's fault. The scene illustrates how researchers' epistemic commitments become embodied by project staff members. The simple exchange of a small amount of money is an act with far-reaching consequences, at least in the eyes of the fieldwork teams, who suggested that paying the man for his lost property was a gesture of good faith and epitomized the project's ethical commitment to do no harm. Giving the money, they said, ensured that the man in question would not go back to his household or village with bad feelings for the project that could influence whether he, his family, or friends welcomed the project in the future or participated in the survey (it was unknown whether this particular man was in the sample). Aside from the formal introductions to district offices, traditional authorities, and others who can influence the tenor of data collection in a sample area, informal, improvised, and tactical social relations directed toward maintaining sample purity and treating the sample with respect played a central role in enabling smooth data collection in the present and the future.

Conclusion

This chapter has emphasized that good data do not lie passively in wait to be collected by fieldworkers. Instead, the shared imaginary of data compels fieldworker and respondent to meet face to face, and clean data are imagined and materialized by standards translated into the field by fieldworkers. In zooming in on some of the hundreds of research encounters that transpire every day in the field, we see that data are cooked and cleaned in multiple stages as they travel to the office or enter a database: raw data, indeed, are an imagined

fiction (Gitelman 2013). Data's value is produced in the frictions that arise when the abstract epistemic investments that define clean data are translated into the particular spaces and embodied social relations of the messy field and in the messy editing practices undertaken by fieldworkers as they handle data before they reach the office. Indeed, the numbers produced are artifacts of the situated negotiations of survey research worlds more than they count or document rural realities.

The pieces of information recorded by fieldworkers like Janet, Tapika, Ishmael, Henry, Collins, and Edward, having subsequently passed through the hands of supervisors and data clerks, are now ensconced in the ordered and sterile space of the database. How do these aggregated data now traverse the boundary between producers and users? How do they reach the audiences who arbitrate their value as evidence for policy or other uses? Chapter 5 traces the next step in data's life course: its re-presentation and ordering in venues ranging from policy meetings to journal articles to conferences.