

6

Descriptive Accuracy in Interview-Based Case Studies

Jennifer Widner

6.1 Introduction

Social scientists and policy-makers care deeply about their ability to draw clear causal inferences from research – and justifiably so. But descriptive accuracy also matters profoundly for the success of this enterprise. Correctly identifying relevant parties, choice points, and perceptions, for example, strongly impacts our ability to understand sources of influence on development outcomes, successfully disrupt and overcome obstacles, and identify scope conditions. The challenge is how to tease out this kind of information in interview-based qualitative research.

This chapter draws on a decade of experience in developing policy implementation case studies under the auspices of a Princeton University program called Innovations for Successful Societies in order to highlight ways to address some of the most common difficulties in achieving descriptive accuracy. The program responded to a need, in the mid-2000s, to enable people leading public sector reform and innovation in low-income and low-middle-income countries to share experiences and evolve practical wisdom suited to context. To develop reasonably accurate portrayals of the reform process and create accurate after-action reports, the program carried out in-depth interviews with decision-makers, their deputies, and the other people with whom they engaged, as well as critics. The research employed intensive conversation with small-N purposive samples of public servants and politicians as a means of data collection.¹ In the eyes of some, this interview-generated

¹ For a compelling argument about the use of nonprobability sampling in elite interview-based case studies, see Tansey (2007).

information was suspect because it was potentially vulnerable to bias or gloss, fading or selective memory, partial knowledge, and the pressures of the moment. Taking these concerns seriously, the program drew on research about the interaction between survey design and respondent behavior and evolved routines to boost the robustness of the interviews it conducted.

6.2 Background

Accuracy is related to reliability, or the degree of confidence we have that the same description would emerge if someone else reviewed the data available or conducted additional interviews.² Our methods have to help us come as close as possible to the “true value” of the actual process used to accomplish something or the perceptions of the people in the room where a decision took place. How do we ensure that statements about processes, decisions, actions, preferences, judgments, and outcomes acquired from interviews closely mirror actual perceptions and choices at the time an event occurred?

In this chapter, I propose that we understand the interview process as an exercise in theory building and theory testing. At the core is a person we ask to play dual roles. On the one hand, our interviewees are a source of facts about an otherwise opaque process. On the other hand, the people we talk to are themselves the object of research. To borrow the words of my colleague Tommaso Pavone, who commented on a draft of this chapter, “The interviewee acts like a pair of reading glasses, allowing us to see an objective reality that is otherwise inaccessible.” At the same time, however, we are also interested in that person’s own perceptions, and subjective interpretations of events, and motivations. “For example,” Pavone said, “we might want to know whether and why a given meeting produced divergent interpretations about its relative collegiality or contentiousness, and we might subsequently probe how the interviewee’s positionality, personality, and preferences might have affected their views.”

In both instances, there are many potential confounding influences that might blur the view. Some threats stem from the *character of the subject matter* – whether it is comparatively simple or causally dense (in Michael Woolcock’s terms, i.e., subject to many sources of influence, interactions, and feedback loops; see Chapter 5, this volume), whether it is socially or politically

² A good description also represents what it is that we aim to study or report. That is, the measures and facts used are logically linked to the key concepts.

sensitive, or whether the person interviewed is still involved in the activity and has friends and relatives whose careers might be affected by the study. Other threats emanate from the nature of *the interviewee's exposure to the subject matter*, such as the amount of time that has elapsed since the events (memory), the extent of contact (knowledge), and the intensity of engagement at the time of the events. Still other influences may stem from the *interview setting*: rapport with the researcher, the order and phrasing of questions, whether there is a risk of being overheard, and the time available to expand on answers, for example.

Our job as case writers is to identify those influences and minimize them in order to get as close as possible to a true account, much as we do in other kinds of social science. And there are potential trade-offs. A source may be biased in an account of the facts, but, as Pavone suggested in his earlier comments, "he may be fascinating if we treat him as a subjective interpreter," whose gloss on a subject may reveal how decision-makers rationalized intense investment in a particular outcome or how they responded to local norms in the way they cloaked discord.

The following section of this chapter treats the pursuit of descriptive accuracy as an endeavor very closely aligned with the logic used in other types of research. Subsequent sections outline practices we can mobilize to ensure a close match between the information drawn from interviews and objective reality – whether of a process or of a perspective.

6.3 The Interview as Social Science Research

An interview usually aims to take the measure of something, and any time a scientist takes a measurement, there is some risk of error, systematic or random. We can try to reduce this problem by modeling the effects of our instruments or methods on the information generated. The goal is to refine the interview process so that it improves the accuracy and completeness of recall, whether of events and facts or of views.

First, let's step back a bit and consider how theory fuels qualitative interviewing, for even though we often talk about this kind of research as inductive, a good interview is rarely free-form. A skilled interviewer always approaches a conversation self-consciously, with a number of key ideas and hypotheses about the subject and the interviewee's relationship to that subject in mind. The inquiry is exploratory, but it has a strong initial deductive framework.

This process proceeds on three dimensions simultaneously, focused at once on the subject matter; the interviewee's preferences, perceptions, and biases; and the need to triangulate among competing accounts.

Theory and interview focus. The capacity to generate good description draws heavily on being able to identify the general, abstract problem or problems at the core of what someone is saying and to quickly ask about the conditions that likely set up this challenge or the options likely tried. At the outset, the interviewer has presumably already thought hard about the general focus of the conversation in the kinds of terms Robert Weiss outlines in his helpful book, *Learning from Strangers*³ – for example, describing a process, learning how someone interpreted an event, presenting a point of view, framing hypotheses and identifying variables, etc. In policy-focused case studies, we begin by identifying the broad outcomes decision-makers sought and then consider hypotheses about the underlying strategic challenges or impediments decision-makers were likely to encounter. Collective action? Coordination across institutions? Alignment of interests or incentives (principal–agent problems)? Critical mass? Coordination of social expectations? Capacity? Risk mitigation? Spoilers? All of the above? Others?

Locking down an initial general understanding – “This is a case of what?” – helps launch the conversation: “As I understand it, you faced ___ challenge in order to achieve the outcomes the program was supposed to generate. How would you characterize that challenge? . . . This problem often has a couple of dimensions [fill in]. In what ways were these important here, or were they not so important?” Asking follow-up questions in order to assess each possible impediment helps overcome problems of omission, deliberate or inadvertent. In this sense, accuracy is partly a function of the richness of the dialogue between the interviewer's theoretical imagination, the questions posed, and the answers received.

The interview then proceeds to document the steps taken to address each component problem, and here, again, theory is helpful. The characterization of the core strategic challenges spawns a set of hypotheses. For example, if the key delivery challenge is the need for collective action, then we know we will have to ask questions to help assess whether the outcome sought was really a public good as well as how the decision-maker helped devise a solution, including (most likely) a way to reduce the costs of contributing to the provision of that public good, a system for monitoring contributions, or whether there was one person or organization with an exceptional stake in

³ Weiss (1994), pp. 9–10.

the outcome and therefore a willingness to bear the costs. In short, the interview script flows in large part from a sense of curiosity informed by theory.

Theory also helps us think about the outcomes we seek to explain in a case study and discuss in an interview. In policy-relevant research we are constantly thinking in terms of measures or indicators, each of which has an imperfect relationship with the overarching development outcome we seek to explain. A skilled interviewer comes to a conversation having thought deeply about possible measures to evaluate the success or failure of an action and asks not only how the speaker, the interviewee, thought about this matter, but also whether any of these plausible measures were mooted, understanding that the public servants or citizens involved in a program may have had entirely different metrics in mind.

Often the outcomes are not that easy to measure. Nonetheless, we want something more concrete than a personal opinion, a thumbs up or thumbs down. To take one example, suppose the aim of a case study is to trace the impact of cabinet management or cabinet design on “ability of factions to work together.” This outcome is not easy to assess. Certainly, we want to know how people themselves define “ability to work together,” and open-ended questions are initially helpful. Instead of nudging the interviewee’s mind down a particular path, the interviewer allows people to organize their own thoughts.⁴ But a skilled interviewer then engages the speaker to try to get a better sense of what it was about people’s perceptions or behavior that really changed, if anything. In the abstract it is possible to think of several possible measures: how long it took to arrange meetings to coordinate joint initiatives, how often requested meetings actually took place, how often the person interviewed was included in subcabinet deliberations that involved the other political parties, whether the person interviewed felt part of the decision process, whether the deputy minister (always from the other party in this instance) followed through on action items within the prescribed timeframe, whether there was name-calling in the meeting room or fistfights, etc.

An interviewer also wants to figure out whether the theory of change that motivated a policy intervention actually generated the outcomes observed. Again, theory plays a role. It helps to come to an interview with plausible alternative explanations in mind. In the example above, maybe power-sharing had little to do with reducing tension among faction leaders.

⁴ See Aberbach and Rockman (2002).

Instead, war weariness, a shift in public opinion, a sharp expansion of economic opportunity outside government that reduced the desire to remain in government, the collapse of factional differences in the wake of demographic change, the personality of the head of government – any of these things might also have accounted for the results. A skilled interviewer devises questions to identify whether any of these causal dynamics were in play and which facts would help us understand the relative importance of one explanation versus the others.

Let me add a caveat at this point. Theoretically informed interviewing leads us to the kinds of descriptive detail important for analysis and understanding. However, it is also crucial to remember that our initial frameworks, if too narrowly defined, can cause us to lose the added value that interviews can generate. As Albert Hirschman (1985) noted years ago, paradigms and hypotheses can become becoming a straightjacket ('confirmation bias'), and the unique contribution of interview-based research is that it can foster a dialogue that corrects misimpressions. Openness to ideas outside the interview script is important for this reason

For example, understanding the source of political will is important in a lot of policy research, but sometimes the most important outcome the lead decision-maker wants to achieve is not the one that most people associated with a policy know about or share. Say we want to use interview-based cases to help identify the conditions that prompt municipal public works programs and other city services to invest in changes that would improve access to early childhood development services. It soon becomes clear that the mayors who had made the most progress in promoting this kind of investment and collaboration sought outcomes that went well beyond boosting children's preparedness for preschool, the initial supposition, and, moreover, each wanted to achieve something quite distinctive. For some, the larger and longer-term aim was to reduce neighborhood violence, while for others the ambition was to diminish inequality or boost social capital and build trust. The open-ended question "Why was this program important to you?" helps leverage this insight.

Theory and the interview process . Interviewing employs theory in a second sense as well. To reveal what really happened, we have to weed out the details people have remembered incorrectly while filling in the details some never knew and others didn't consider important, didn't want to highlight, or simply forgot. Therefore, in the context of an interview, it is the researcher's job not only to seek relevant detail about processes, but also to perceive the

gaps and silences and use additional follow-up questions or return interviews to secure explanations or elaboration.

In this instance, the researcher navigates through a series of hypotheses about the speaker's relationship to the issue at hand and knowledge of events. A few of the questions that leverage information for assessing or weighting answers include:

- “You had a peripheral role in the early stages of these deliberations/this implementation process, as I understand it. How did you learn about the rationale for these decisions? [Co-workers on the committee? Friends? Briefed by the person who led the committee or kept the minutes? Gleaned this information as you began to participate?] How would you say that joining the deliberation/process/negotiation late colored your view of the issues and shaped your actions, or did it not make much difference?”
- “You were involved in a lot of difficult decisions at the time. How closely involved were you in this matter? Did you spend a lot of time on it? Was it an especially high priority for you, or was it just part of your daily work?” “Given all the difficult matters you had to deal with at the time, how greatly did this issue stand out, or is it hard to remember?” (Level of knowledge helps the interviewer weight the account when trying to integrate it with other information.)
- “The other people involved in this decision/process/negotiation had strong ties to political factions. At least some of them must have tried to influence you. At what stages and in what form did these kinds of pressures arise?” “Were some voices stronger than others?” “How would you say these lobbying efforts affected your decision/work/stance?”
- “As I understand it, you took a decision/action that was unusual/worked against your personal interest/was sure to be unpopular with some important people. How would you characterize your reasons for doing so?”

The information these questions leverage helps the case writer assess the likely accuracy of an account, in at least three ways: First, it helps us understand whether someone was in a position to know or heard about an action secondhand. Second, it helps us assess the integrity of a response – for example, does a statement run contrary to the speaker's obvious personal interests and is it therefore more believable? Third, it can also help spot purely opportunistic spin: Is the view expressed consistent with the speaker's

other actions and attitudes, or is at odds with these? Can the person offer a clear story about this divergence, or did this perception or preference evolve in association with a promotion, an election, or some other event that may have influenced behavior?

Theory and ability to arbitrate among competing statements. The third task of interview-based case research is to meld information garnered from different conversations and other types of sources in order to triangulate to the truth. This too is a theory-driven enterprise. Every time there is a clash between two assertions, we ask ourselves the familiar refrain “what could be going on here?” (hypothesis formation, drawn from theory), “how would I know?” (observable measures), and “by what method can I get this information?” (framing a follow-up question, checking a news report, consulting a register, etc.).

We may weigh the account of someone who joined a process late less heavily if it clashes with the information provided by those closer to a process, but it could be that the latecomer is less vulnerable to groupthink, has no reputation at stake, and offers a clearheaded story. Maybe we know the person was brought in as a troubleshooter and carried out a careful review of program data, or that the person is highly ambitious, eager to appear the hero who saved a failing initiative that, in fact, had not performed as badly as stated? Career paths, reputational information, and the written record – for example, longitudinal performance data – can all assist in making sense of disparate accounts.

This thought process may have to take place in the context of an interview as we listen and form follow-up questions, but it can also fuel exit interviews or second conversations designed both to provide another occasion to relate events remembered after the first encounter and to afford a chance to react to divergent information or ideas others may have voiced. This is the task of the exit interview.

To stress that skilled interviewing is theory-driven does not mean social scientists do a better job than journalists. Journalists might call the same kind of thought process “intuition” or “savvy,” but, when asked to step back, be self-conscious, and break down the mental exercise involved, the reality of what they do differs little from how social scientists or historians proceed in their work. The editor who tells a cub reporter, “I smell a rat” upon hearing the sketch of a story is positing an alternative hypothesis to test the adequacy of a description. Employing a general model built on experience, the editor pushes the reporter to use evidence to identify what

the people described in the reporter's draft are really doing.⁵ A reporter's intuition is equivalent to a social scientist's skill in quickly framing plausible hypotheses and crafting a follow-up question that will yield the evidence to arbitrate among conflicting accounts – conducting social science inquiry “on the fly.”

Regardless of the interviewer's own background, skilled interviewing places a premium on preparation. Even if the interviewer does not have a full-blown research design, it is crucial to have a preconversation sketch that frames hypotheses about the subject matter of the case and alternative plausible explanations; specifies the role the interviewee played in these events and the level of knowledge or type of gloss that relationship might produce (or at least does so with as much care as possible at this stage); and summarizes what archival sources say about the story line. This practice helps frame the questions that will tease out the evidence that disconfirms, modifies, or corroborates different versions of a story mid-conversation as new facts and observations emerge.

6.4 Improving Recall and Specificity

Solid, substantive preparation alone does not generate the requisite level of detail and accuracy needed in a policy case study. The skilled interviewer also has to overcome barriers to cognition. The people we interview are busy. They often work in a language different from ours. They may not understand what a study is about and what kinds of information the interviewer seeks. Further, like the rest of us, they forget and they tire. As a result, their answers to questions may vary from one interview to the next, making the descriptions we assemble less reliable.

Survey researchers have struggled with these challenges for decades.⁶ They have investigated how people answer questions and how to improve accuracy in responses. Their reflections are helpful for those who do qualitative interviews.

⁵ Although we are often insufficiently self-conscious about this thought process, the practice is broad and has long-standing roots. Just think of the common English phrase “I smell a rat,” which signals that the speaker has an alternative, probably more cynical, hypothesis about something a conversational partner has just described than the partner has offered. The origins of the phrase go back to the 1500s. Samuel Johnson defined its meaning as “to be put on the watch by suspicion as the cat by the scent of a rat; to suspect danger.” All is not as it seems: Johnson (1755).

⁶ For example, Cannell, Miller, and Oksenberg (1981); Wright, Gaskell, and O’Muircheartaigh (1994); and Krosnick (1991).

1. One fairly obvious starting point or maxim is to make sure that the interviewee understands the purpose of the project or study and can perceive the level of detail expected. It is common for someone to ask quizzically, “Why would anyone be interested in what I did?”

Helping a speaker understand the intended audience improves motivation and accuracy. With respect to policy-focused interviews, asking someone to help a peer learn from a program’s experience usually changes an interviewee’s mental stance, and enables the person to hone in on the kind of subject matter sought and the level of operational detail needed. In the Princeton program we often used the phrase, “The purpose is to help your counterparts in other countries learn from your experience” in the invitation letter and follow-up, as well as in the lead-in to the interview itself. We also emphasized that “the aim of the case study is to profile the reform you helped design, the steps you took to implement the new system, and the results you have observed so that others can learn from you.” Periodically, we reiterated these points. When an interviewee can imagine a conversation with the person who will use the information, answers are more likely to be specific. It also becomes easier to induce someone to be compassionate and speak honestly about the real problems that arose during a process, so that the target group of readers don’t go astray or fail to benefit from the experience.

2. A second maxim is to ensure questions are clear so the interviewee does not have to struggle with meaning. A long, rambling question that requires energy to parse can sink an interview. By contrast, a simple, open-ended “grand tour” question is often a good place to begin, because many people are natural storytellers, become engaged, and start to focus their comments themselves when given this latitude. In his ethnographic interview classic, for example, Spradely suggests asking “Could you describe what happened that day?” or “Could you tell me how this office works?” Subsequent questions can focus on the elements of special relevance to the subject and may include prompts to reach specific subject matter or the requisite level of detail.⁷

⁷ Spradely (1979). The people who train investigators or prosecutors to conduct interviews have experimented with interview formats. “Structured interviews” usually begin with an open question and then focus on particular points raised. Instructions to “remember back to ____” can trigger improved recall as well. Adding “yes/no” questions can generate additional information (and sometimes aide in detecting prevarication).

3. In framing questions, we try to avoid ambiguous or culturally loaded terms that increase the amount of mental calculation an answer requires.⁸ How much is “usually” or “regularly”? “Big”? How many years is “young” or “recently”? It may be better to ask, “About how often did that happen during that year?” or “How many times did that happen that year?” Novice interviewers often refer to seasons as they pinpoint the time of an action, but of course these vary globally, so the references merely confuse (moral: benchmark to events or national holidays).

Similarly, we try to eliminate questions that require the interviewee to talk about two different things – the “double-barreled question” or compound question. “Did that step reduce error?” is clear, but “Did that process reduce error and delay?” asks about two dimensions that may not be related, yet it seems to require one answer. In this instance, it does not take much effort to sort out the two dimensions and in an interview context, as opposed to a survey, that is feasible. However, a speaker will have a slightly tougher time with a compound question about a preference, motivation, or interaction: “Was the main challenge to compensate those who would have to alter their farming practices and to help the community monitor illegal deforestation?” “Was this group an obstacle to winning the vote in the legislature and a source of public backlash?” Simple questions and quick follow-ups usually elicit better information than complex questions that ask for views on two or more things at once.

4. The passage of time influences the ability to remember information and potentially also makes it hard to check the reliability of a description. In the 1980s, studies of physician recall found that memory of specific patient visits decayed very rapidly, within two weeks of a visit.⁹ Norma Bradburn and her colleagues reported that about 20 percent of critical details are irretrievable by interviewees after a year and 60 percent are irretrievable after 5 years.¹⁰ The ability to remember distant events interacts with the salience or importance of the events to the interviewee and with social desirability. A well-received achievement that occurred two years earlier may be easier to remember than something that did not work very well or was not considered important at the time. Using “probes,” or questions that fill in a little detail from archival research, can help break the mental logjam.

Phrasing that takes the interviewee carefully back in time and provides reminders of the events that occurred or the locations in which they occurred

⁸ Wright, Gaskell, and O’Muircheartaigh (1994). ⁹ Cannell, Miller, and Oksenberg (1981) p. 397.

¹⁰ Bradburn, Rips, and Shevell (1986) p. 158.

may improve recall. Specific dates rarely have the same effect (imagine trying to remember what you were doing in August three years ago). Recall can improve during the course of an interview or after the interviewer has left.

The passage of time may also alter perceptions. Views change, and the interviewee may subconsciously try to harmonize interests, attitudes, or opinions. As in a historical account, what the principal actors knew at the time they recognized a problem and decided what to do is very important to capture accurately, and it may take some extra effort to trigger memory of initial perceptions and how these changed. Here is an example from a series of cases on the 2014 West Africa Ebola Outbreak Response.¹¹

Example: Effects of time on accuracy (two interviews conducted in late 2015 about the Ebola response):

Interview 1: Question: “How useful was the US military response to improving logistics capability?” “The US timing was all wrong. The military built emergency treatment centers that were never used because the epidemic ended by the time the centers were ready. The US military action was irrelevant.”

Interview 2: Question: “Let’s go back to August and September 2014 when the outbreak escalated dramatically in Liberia. Could you talk about the impact of the US military on logistics?” “In September 2014, the models said the number of people infected would rise to over a million. The US military prepared for that eventuality. Later the epidemic declined and the ETUs [emergency treatment units] weren’t used, but in the end what seemed to matter to the public was the visible sign that a big power cared, which generated a psychological boost. We hoped the military would be more useful in moving lab materials around but they had instructions not to enter areas where an outbreak had occurred so they just dropped us at the edges of these areas and then we made our way from there.”

There is some truth to both statements but the timestamp in the second question elicited a more complete answer that helped resolve tensions among accounts.

5. Memory of actions taken in a crisis atmosphere, when people may have worked intensely on many different fronts, tends to be less good, emerges in a highly fragmented form with high levels of error, or acquires a gloss. Said one ISS interviewee who had worked intensely on a disaster response, “As we talk, I can feel PTSD [post-traumatic stress disorder] coming back.” Words tumbled out, and the interviewer had to piece together the order in which actions occurred.

In these circumstances, it is helpful to plan one or more return interviews. Between sessions, people will tend to remember more, though their

¹¹ See Widner (2019a and 2019b).

memories may also start to embellish or spin the account. Questions that contain specific information about the circumstances and ask for a reaction may help alleviate that problem. For the researcher, the challenge then becomes integrating the different versions of an event to ensure that they synchronize accurately.

6. Research on how respondents react to surveys suggests that question order can make a big difference in the responses people offer.¹² Although there is not parallel research on long-form qualitative interviews, it stands to reason that some of the same issues arise in this slightly different context, although it is easier for the interviewer to circle back and ask for views a second time than it might be in a survey, providing a possible corrective.

In designing and modifying the informal script that structures the interviews, it may help to consider how the sequence or juxtaposition of particular questions might influence what people say by inadvertently priming a particular response. For example, if one question focuses attention on the influence an interest group brought to bear on a decision, the answer to an unrelated question may place heavier emphasis on interest groups than it would have in a different question lineup. Sometimes the best cure for this type of spillover is to acknowledge it directly: “We have talked a lot about interest group influence. I want to change the topic and focus on ____, now. Although there may have been some interest group influence, most likely other things were important too, so I encourage you to step back and think about what shaped this decision, more broadly.” An alternative is to shift to a different, more minor topic – or recommend a brief break – before returning to the line of questioning.

In policy research, political or social sensitivity may lead to self-censorship. To lessen this response, while also respecting the risks a speaker faces, it is sometimes possible to sequence questions so that they enable the speaker to articulate a problem in a diplomatic way, threading the needle: “I imagine that people who had invested in that land were upset that the city wanted to build a road there. Did any of those people ever speak about this problem in public? Did any of them ever come here to express their views? I see in the newspapers that politician X owned some land in that area – was he part of the group that objected? Did the program change after this point?”

Pacing sensitive questions may necessitate extra care in order to prevent the interviewee from calling an end to the conversation or from shifting to highly abbreviated responses. If the point of an interview is to acquire

¹² For example, see Zaller and Feldman (1992) and Schwarz, Oyserman, and Peytcheva (2010).

information about a particular stage of a negotiation, then it may be better to proceed to that point in the conversation before posing sensitive questions about earlier matters – and then loop back to these other sensitive issues when asking about results or reflections toward the conclusion of the conversation. By that point the interviewer has had a chance to build credibility and signal facility with some of the technical details, making it more likely the speaker feels s/he has had a fair chance to explain actions and views, while also realizing the interviewer is unlikely to be satisfied with a stock answer. Ethics rules that require returning to the speaker for permission to use quotes or identifying information can also assist willingness to speak, provided the speaker trusts that the interviewer will indeed live up to this commitment. (Note that this research ethics commitment runs counter to the standards in journalism, where the emphasis is on conveying publicly important information in real time and not allowing the holders of that information to act as censors.)

Ending on a more positive note is also helpful, both for the well-being of the interviewee and for maintaining the goodwill that makes it possible to return to the conversation later: “You accomplished __, __, and ____. When you think back on this episode are there other things that make you especially proud/happy/satisfied with the work/____?”

7. Offer the right kinds of rewards. Because it takes a lot of mental energy to respond to questions and because there is no immediate tangible reward, an interview has to generate and sustain motivation. Usually, helping someone understand the important purpose and specific focus increases interest. Most people also want some sense that they are responding with useful information. If they don’t have this sense, they will drop out.

There is a fine line between leading, on the one hand, and nondirective feedback that merely sustains a conversation. A leading question suggests correct answers or reveals the researcher’s point of view. This type of feedback reduces accuracy. By contrast, there are neutral forms of feedback that can motivate and lead the interviewee to persist in answering questions. Cannell, Miller, and Oksenberg (1981: 409–411) suggest a set of four responses that ISS has also found helpful:

- “Thanks, that’s useful, OK.”
- “I see. This is the kind of information we want.”
- “Thanks, you have mentioned __ things . . .”
- “Thanks, we are interested in details like these . . .”

Speakers often model the length of their responses on the interviewer's behavior. Rewarding specificity in a response to an open-ended question early in the interview with "Thanks, we are interested in details like these" can send the right signal (assuming detail is in fact what we want).

6.5 Integrating Streams of Evidence, Arbitrating Differences

In survey research, social scientists aggregate data from multiple respondents by analyzing central tendencies, assessing variance, and then evaluating the influence of causal factors on responses to questions using some type of regression analysis. Although less concerned with central tendencies and average effects, qualitative case study research also has to integrate multiple streams of information from interviews – information about views as well as processes. This stage of the research can catch and reconcile discrepancies or spin, but it can also become a source of error if the researcher incorrectly judges one account to be more truthful than another, with little basis in fact or little transparency about the reasons for privileging a particular point of view.

Arbitrating among conflicting streams of evidence takes place in journalism every day, and the adages journalists follow are equally applicable in social science research. Editors and reporters term the failure to resolve a contradiction or a clash of perspectives as "he said, she said" journalism.¹³ Columbia University journalism professor Jay Rosen, who led the charge against "he said, she said" reporting, offered an illustration. In this instance, a US National Public Radio reporter described a controversy over new reproductive health regulations and said that one group portrayed the rules as "common sense" while another saw them as designed to drive clinics out of business. The reporter laid out each group's claims and moved on. Rosen cried foul and said the reporter had an obligation to offer a more complete description that gave the reader some sense of the evidence underlying the seemingly disparate claims. This imperative has grown stronger as quality journalism has tried to combat disinformation.

Rosen's remedies were exactly those his social science counterparts would have offered: hypothesis formation, measurement, and follow-up questions. In this instance, Rosen said, the reporter could have compared the new regulations to those already in place for similar procedures in the same state and to regulations in other jurisdictions so the reader could see whether

¹³ Rosen (2011).

the claim that the new rules were “common sense” had some basis in fact. The reporter could have read the safety report to see whether the accident or infection rates were especially high compared to related procedures. In short, Rosen argues, the researcher’s obligation to the reader is to resolve discrepancies when they involve matters that affect the reader’s ability to make a judgment about the core subject matter of the case. The reader’s mind should not buzz with further questions, and the description must have all the components necessary to understand the intervention described, including those an expert would consider fundamental.

Discrepancies in streams of interview evidence can arise from many sources, including differences regarding when two people became involved in a process, the roles they played and the knowledge available to them in each of these roles, and the life experiences or technical skills they brought to the job. That is, disagreements do not always arise from deliberate spin. Here are three examples of descriptive or reporting challenges drawn from ISS case study research and the intellectual process these challenges triggered.

One: Superficially discrepant timelines (case about the Liberian Ebola Outbreak response coordination¹⁴):

Question: When did Liberia adopt an incident management system for responding to the Ebola outbreak?

Interview 1: CDC Director Tom Frieden persuaded President Ellen Sirleaf to support an incident management system for coordinating the Ebola response. (From archival record: This meeting took place on or around August 24, 2014, on Frieden’s visit to the country.)

Interview 2: A CDC team visited Monrovia the third week in July 2014 and began to work with officials to set up an incident management system. (From archival record: The president appointed Tolbert Nyenswah head of the incident management system on August 10.)

Thought Process: The interviewer seeks accuracy in describing a sequence of events. At first blush it might seem that one subject just remembered a date incorrectly, but the archival evidence suggests that the dates of the events cited are indeed different. What else could be going on? One hypothesis is that something happened in between the two periods that required the president to revisit the choice of approach. An interviewer in strong command of the timeline might then frame a follow-up for interviewee 1: “Could you clarify the situation for me? I thought that the president had earlier appointed someone to head an incident management system. Did the first

¹⁴ Widner (2019a and 2019b).

effort to launch the system fail or flounder?” For interviewee 2: “I understand that later the president and the head of the CDC discussed whether to continue the system in late August. Did anything happen in mid-August to shake the president’s confidence that the IMS was the right approach?”

Two: Superficially discrepant information about states of mind or relationships: (case on cabinet coordination in a power-sharing government¹⁵)

Interview 1 (with someone who was in the meeting): “The dialogue process helped resolve stalemates and we emerged from these sessions in a better position to work together.”

Interview 2 (with a knowledgeable observer who was not in the meeting): “The dialogue process just helped the parties delay taking steps to meet the goals they had jointly agreed to. The leaders argued for long periods.”

Thought Process: In this instance, the researcher wants to know whether tensions among political parties in a unity government were lower, about the same, or higher after resort to an externally mediated “dialogue” mechanism. There are three challenges. First, few people were in the room and the perceptions may vary with knowledge. Second, “tension” or “trust” among political parties is something that is “latent” or hard to measure. Third, delay and levels of distrust could be related in a wide variety of ways. Delay might have increased trust or decreased it.

In this instance, the researcher would likely have to return to the people interviewed with follow-up questions. One might venture several hypotheses and ask what evidence would allow us to rule out each one, then frame questions accordingly. Did the number of matters referred to mediation go down over time? Did the number of days of mediation required diminish over time? Did deputy ministers perceive that it became easier or harder to work with colleagues from the other party during this period? Did progress toward pre-agreed priorities stall or proceed during this period?

If what went on in the mediation room is confidential, then the researcher has to frame questions that rely on other types of information: “Comparing the period before the mediation with the weeks after the mediation, would you say that you had more purely social conversations with people in the opposite party, fewer, or about the same? Was there a new practice introduced after the mediation that affected your ability to have these conversations?”

Three: Insufficient detail; “the mind does not come to rest” and the reader is left with an obvious, big, unanswered question. This challenge arises frequently. For example, the Princeton ISS program ran into this issue in trying to document the introduction of a public service delivery tracking system in the Dominican Republic.¹⁶

¹⁵ Schreiber (2016). ¹⁶ Cameron (2016).

Interview 1 (with an officer responsible for tracking action items in a ministry): “At first we added data to the tracking system each month but after a few months everything slowed down and we added information only every three or four months.”

Interview 2 (with an officer responsible for overseeing the central recording process): “Some ministries didn’t report at all. They never added information to the tracking system.”

Thought process: There is a discrepancy between the two statements, but in both instances it is clear that work had ground to a halt. The issue is when and why. Was the new system unworkable in all ministries or just some? Further, was the system impossible for most to use, or was there something else going on? One could ask a general question, “Why did that happen?” – an approach that often yields surprising answers. But hypotheses and follow-ups might also help winnow out plausible from less-plausible explanations. Did a few ministries try to report and then give up and join the others in noncompliance, or did a few continue to report? Then to the rationale: Was there no progress to report? Was there no penalty for not reporting? Was it hard to find the time to file the report? Did the software break down, or was there limited electrical power to run the system? Was someone designated to acquire and upload the information, or was no one really in charge of that function? Were the instructions hard to follow? Did the minister care or say anything when reporting slowed or halted? Did anyone from the president’s office/delivery unit call to ask why the report was slow? Was there pressure to delay the reports? Why?

If there are no data available to resolve a contradiction or settle a logical subsidiary question, then the researcher can say so. Andrew Bennett has proposed valuing evidence from different interviews according to kind of schedule of plausibility.¹⁷ Attach an estimate or “prior” to the possible motives of each interviewee who provides evidence and weigh the evidence provided accordingly. If someone offers evidence that clearly does not make that person “look good,” one might have more confidence in the other information offered. “Social psychologists have long noted that audiences find an individual more convincing when that person espouses a view that is seemingly contrary to his or her instrumental goals,” Bennett suggests; “For similar reasons, researchers should follow established advice on considering issues of context and authorship in assessing evidence. Spontaneous statements have a different evidentiary status from prepared remarks. Public

¹⁷ Bennett and Checkel (2015: 24–25).

statements have a different evidentiary status from private ones or from those that will remain classified for a period of time.”

6.6 Selection Bias

The protocols used to select interviewees are always a potential source of error, whether in survey research or in qualitative interviewing. In process-tracing case studies, we choose the people to interview instrumentally. That is, we want information from people who have direct knowledge of a process. But there is a consequent risk of interviewing only people from a single political party; from the group tasked with the daily work of carrying out a program; or from one ethnic, religious, or economic group affected. This problem may not always be damning. It may be partly contingent on the question asked: for example, there may be circumstances when the only people who are in a position to know about a series of actions are indeed people from the same small group. However, if the aim is to know how others perceived a decision or a program, or whether people thought a process was representative – or whether beneficiaries viewed the program in the same way policy-makers did – it goes without saying that we need the views of a broader group of people.

Avoiding selection bias in interview-based research can prove challenging, especially in less open societies. At ISS, which has focused on governmental reform, researchers typically spend much of their time securing an accurate description of a change in structure or practice and its implementation. Usually the only people with that information are those in government who actually carried out the daily legwork. In some settings, these people are likely to have a party affiliation and come only from one political party. They also may not know how the “clients” – the country’s citizens – view what they do. Because of this, the research program made it standard practice to include in its interview lists:

- people most likely to be critical of the reforms we profile (we try to identify such people by looking at local newspaper editorials and headlines, speaking on background with journalists, etc.)
- counterparts from another political party where these exist (predecessors in the same role, for example)
- civic leaders or university researchers who work closely with the intended beneficiaries or clients

- public servants who worked on the project in different locations
- the “victors” and the “vanquished” – the people whose views prevailed and those whose views did not.

Where there are few civic groups it may be particularly difficult to identify people who are close to the views of clients and users and can generalize reliably about perceptions and experiences.

Problem: Critics won't speak (case study on extension of civilian oversight of the military)

Interview 1: The defense minister in the political party that recently came to power says, “We retrenched several thousand soldiers and gave them severance pay. There was no serious objection to the new policy. Some of the senior military officers believed the policy was the right thing to do and supported it.”

Missing Interview: From archival sources we know that a political party led a protest against this very policy. Neither the officers of that political party nor the identifiable leaders of the protest assented to an interview, however.

Remedy: There are some partial solutions for countering selection bias that arises from this sort of “missing actor” problem. One is to try to induce people who will speak on the record to be self-reflective. For example, Jeffrey Berry suggests that the researcher “ask the subject to critique his own case – Why aren't Democrats buying this?” or to say, “I'm confused on one point; I read . . .”¹⁸ Another approach is to draw on the publications the critics have authored. These may not get at the real reasons for the criticisms the groups raise, but they may provide enough information to represent the view, and enough detail for the researcher to use to seek a reaction from those who will go on the record.

Another kind of selection bias can arise in new democracies or more authoritarian political systems: self-censorship. In this situation, because of concerns about vulnerability or because of traditions that discourage openly critical comments, everyone interviewed offers a “careful” or biased response. The question is how to break through the reserve without jeopardizing any participant's safety. One possibility is to identify fractures within the political party – a committee or wing or leadership group that genuinely wants to know how well something works and is willing to talk about suspected problems. We can then use these to frame questions that don't require an interviewee to criticize but instead just ask for steps taken when X happened. Phrasing questions so that they don't force one person to impugn another can also help: “If you had to do this over again, what would you do

¹⁸ Berry (2002), p. 680.

differently?” “If you could advise your counterpart in another country how to do _____, what special advice or tips would you want to convey?”

The act of “getting in the door” can create selection bias problems too.¹⁹ The first people to respond favorably to requests to interviews may be people who are distinctive in some way – those who feel empowered, younger people, people who have aspirations in electoral politics and see a way to lend some credibility to their campaigns. To guard against this kind of bias, it is important to step back periodically and to ensure that the list of those who responded favorably to interview invitations includes people who were involved with what we seek to document but don’t have the same profile as others.

6.7 Conclusion

Qualitative case studies are a form of empirical research. Facts are the currency in which they trade. As such they are potentially vulnerable to the same kinds of problems that bedevil quantitative research, from low measurement validity to data collection techniques that bias the views or accounts surveyed or introduce error. This chapter offers a schema for thinking about these challenges in the context of preparing interview-based process-tracing case studies, along with a few partial solutions to some common problems.

One implication of the observations offered here is that careful interview preparation yields a high return with respect to the accuracy and completeness of a process-tracing case study. That means 1) knowing the subject well enough to frame thoughtful hypotheses and measures in advance, and to build these into draft questions; 2) establishing a timeline and “prestory” from news sources, operations reports, or preliminary “informant” interviews; 3) learning about the backgrounds of the people central to the policy initiative; 4) identifying representatives of the beneficiary groups as well as likely critics or people who had special vantage points; and 5) understanding options tried in other, similar settings or in other periods. This background preparation then shapes the development of interview scripts, useful for thinking hard about clarity, narrative flow, question sequence, and other matters that impinge on the quality of the information elicited. Although the interview itself is a conversation, not (usually) a series of survey questions read off a schedule, the development of the written script sharpens the

¹⁹ Goldstein (2002), p. 669.

interviewer's ability to elicit the information required, while maintaining a positive rapport with the speaker.

A second implication is that we have to be transparent about the basis for arbitrating differences that emerge across interviews. Sometimes the prose can say what the reader needs to know: "Staff members involved at the beginning of the project, when the initial pilot failed, remembered that _____. But those who joined later, after the first results of the revised program started to emerge and neighborhood resistance had dissipated, had a different view of the challenges the government faced." In other instances, a discursive footnote of the sort that Andrew Moravcsik (Chapter 8, this volume) proposes may be the best way to help the reader understand the judgments the author made.

A third implication of this analysis is that the purported differences in ability to rely on quantitatively analyzed survey data, on the one hand, and qualitative interview data, on the other, are vastly overstated. The main difference between the two has more to do with whether frequencies or distribution of perspectives across populations matter to the aim of the project. If they do, then survey data may have greater value. But if the aim is to elicit understanding of strategic interaction or a process, then purposive interviewing will tell us more. In both contexts, however, the same concerns about eliciting accurate responses apply and some of the same remedies prove useful.

References

- Aberbach, J. and Rockman, B. (2002) "Conducting and coding elite interviews," *PS: Political Science and Politics*, 35 (4), 673–676.
- Bennett, A. and Checkel, J. T. eds. (2015) *Process tracing: From metaphor to analytic tool*. New York: Cambridge University Press.
- Berry, J. (2002) "Validity and reliability issues in elite interviewing," *PS: Political Science & Politics*. 35 (4), 679–682.
- Bradburn, N. M., Rips, L. J., and Shevell, S. K. (1986) "Answering autobiographical questions: The impact of memory and inference on surveys," *Science*, 236, 158.
- Cameron, B. (2016) "Delivering on promises: The Presidential Goals System in the Dominican Republic, 2012–2016," *Innovations for Successful Societies*, Princeton University, accessed December 13, 2021 at <https://successfulsocieties.princeton.edu/publications/delivering-promises-presidential-goals-system-dominican-republic-2012%E2%80%932016>.
- Cannell, C., Miller, P., and Oksenberg, L. (1981) "Research on interviewing techniques" in Leinhardt, S. ed., *Sociological methodology*. San Francisco: Jossey-Bass, pp. 389–437.
- Gerring, J. (2012) "Mere description," *British Journal of Political Science*, 42 (4), 721–746.

- Goldstein, K. (2002) "Getting in the door: Sampling and completing elite interviews," *PS: Political Science & Politics*, 35 (4), 669–672.
- Hirschman, A. O. (1985) *A bias for hope: Essays on development and Latin America*. Boulder: Westview Press.
- Johnson, S. (1755) "Smell a rat." In *A dictionary of the English language: A digital edition of the 1755 classic by Samuel Johnson*. Edited by B. Besalke. Last modified: March 26, 2012. <https://johnsonsdictionaryonline.com/smell-a-rat/>.
- Krosnick, J. (1991) "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Applied Cognitive Psychology*, 5, 213–236.
- Leech, B. (2002) "Asking questions: Techniques for semistructured interviews," *PS: Political Science and Politics*, 35 (4), 666.
- Rosen, J. (2011) "We have no idea who's right: Criticizing 'he said, she said' journalism at NPR." Accessed December 13, 2021 at <https://pressthink.org/2011/09/we-have-no-idea-whos-right-criticizing-he-said-she-said-journalism-at-NPR/>.
- Schreiber, Leon. "Making power sharing work: Kenya's Grand Coalition Cabinet, 2008–2013," *Innovations for Successful Societies*, Princeton University, 2016 accessed December 13, 2021 at <https://successfultsocieties.princeton.edu/publications/kenya-powersharing-cabinet>.
- Schwarz, N., Oyserma, D., and Peytcheva, E. (2010) "Cognition, communication, and culture: Implications for the survey response process," in *Survey methods in multinational, multi-regional, and multicultural contexts*, edited by J. A. Harkness et al., John Wiley & Sons, Inc., pp. 175–190.
- Spradley, J. (1979) *The ethnographic interview*. New York: Holt, Rinehart and Winston.
- Tansey, O. (2007) "Process tracing and elite interviewing: A case for non-probability sampling," *PSOnline*, pp. 765–772. Accessed December 13, 2021 at www.cambridge.org/core/journals/ps-political-science-and-politics/article/process-tracing-and-elite-interviewing-a-case-for-nonprobability-sampling/8EE25765F4BF94599E7FBD996CBFDE74.
- Weiss, R. S. (1994) *Learning from strangers: The art and method of qualitative interview studies*. New York: Free Press.
- Widner, J. (2019a) "All hands on deck: The US response to West Africa's Ebola crisis, 2014–2015," *Innovations for Successful Societies*, Princeton University, accessed December 13, 2021 at https://successfultsocieties.princeton.edu/sites/successfultsocieties/files/JW_Ebola_USResponse_Final_June%2028%202018_JRG_0-3_1.pdf.
- Widner, J. (2019b) *Responding to global health crises: Lessons from the US response to the 2014–2016 West Africa Ebola outbreak*. IBM Center for the Business of Government, accessed December 13, 2021, at www.businessofgovernment.org/report/responding-global-health-crises-lessons-us-response-2014-2016-west-africa-ebola-outbreak.
- Wright, D., Gaskell, G., and O'Muircheartaigh, C. (1994) "How much is 'quite a bit'? Mapping between numerical values and vague quantifiers." *Applied Cognitive Psychology*, 8, 479–496.
- Zaller, J. and Feldman, S. (1992) "A simple theory of the survey response: Answering questions versus revealing preferences," *American Journal of Political Science*, 36 (3), 579–616.