Original Article

Aligning Interviewing with Process Tracing[1]

Ezequiel Gonzalez-Ocantos[1] 🅸🅳 and Juan Masullo[2]

Abstract

Interviews play a pivotal role in process tracing (PT) by allowing researchers to delve deep into the intricacies of agency, inter-agent interactions and relationships, and the processes underlying meaning and decision-making. These dimensions are essential for evaluating process theories connecting causes to outcomes in specific cases. Testing theoretical arguments via PT bears implications for how we conceive interviewing. We provide recommendations for scholars to design interview research aligned with PT best practices, focusing on sampling and the design of interview protocols, and being sensitive to differences between PT approaches. Aligning interviews with PT's specific requirements strengthens the weight and inferential power of evidence. While the methodological foundations of PT and related data analysis techniques are well-documented in the literature, there is still a gap concerning data collection and generation. We aim to address this by encouraging process tracers to think systematically about their interviewing plans at the design stage.

Keywords

process tracing, interviewing, qualitative methods, sampling, questionnaire design

[1]Department of Politics and International Relations, Nuffield College, University of Oxford, Oxford, UK
[2]Institute of Political Science, Leiden University, Den Haag, The Netherlands

Corresponding Author:
Juan Masullo, Institute of Political Science. Leiden University, Turfmarkt 99, 2511 DP Den Haag, Netherlands.
Email: j.masullo@fsw.leidenuniv.nl

Introduction

In recent years, we have witnessed the consolidation of standards of best practice when using process tracing (PT) to make descriptive and causal inferences about social and political phenomena based on within-case evidence.[2] PT entails the analysis of large amounts of evidence about context, mechanisms, and sequences. It requires diving into a case to document the existence of an uninterrupted chain of events, motivations, actions, and reactions that together bring about an outcome of interest. While initially scholars tended to equate PT with the "historian's method," emphasizing archival and documentary research,[3] interview research now features prominently among the data-gathering techniques used in PT work.[4]

Interviewing is a powerful technique for collecting and generating data for PT because it allows researchers to delve deep into agency, relations between agents, and individual and collective processes of meaning and decision-making. While not free of challenges,[5] interviews allow researchers to explore deeply how people see themselves and their circumstances (perceptions), how they interpret what they see, say, or do (meaning), and what they say are the reasons behind their actions (motivation) (Small and Calarco 2022; Small and Cook 2023). For decades, these aspects have captured the attention of both sociologists and political scientists and are often at the core of the processes that process tracers aim to unpack. Why do Syrian refugees living in refugee camps mobilize against authority in Jordan and not Turkey or Lebanon (Clarke 2018)? Why do some parties formed by social movements end up developing highly hierarchical structures (Anria 2018)? Why did rural peasants support leftist guerrillas during El Salvador's civil war, taking mortal risks and with no real expectations of material gain (Wood 2003)? These are all outcome(y)-oriented or "causes-of-effects" questions (Gerring 2007; Goertz and Mahoney 2012, Chapter 3) that have been addressed via PT and with interviews as a chief data-collection technique.[6] Yet, while there is a large and rich literature on how to master the art of interviewing in political science and especially in sociology,[7] little has been written specifically about it when the goal is to probe theoretical arguments via PT.[8]

We contend that testing theoretical arguments via PT bears implications for how we think of and practice interviewing. We develop a series of recommendations for how scholars can design interview research to generate data that is particularly useful for testing and refining theories about processes linking a cause (or set of causes) to an outcome and that comply with PT "best practices." Concretely, we focus on two aspects of interview research:

(a) whom we choose to interview—*sampling*; and (b) how we structure questionnaires—*designing interview protocols*. After briefly discussing why interviews are valuable for process tracers, the article first tackles issues related to sampling. In 2021, an important Sociology journal, Gender & Society, amended its submission guidelines to warn prospective contributors that "papers with exclusively interview data and fewer than 35 cases [interviewees] are scrutinized carefully." The editors' assumption seems to be that when it comes to interview-based research, there is safety and quality in numbers. Should one, therefore, automatically conclude that of two articles using PT and interviews recently published in a leading Political Science journal, World Politics, the evidentiary weight of the one based on over 600 interviews (Idler 2020) is superior to the one that "only" features 35 (Closa and Palestini 2018)? The answer, we argue here, is no. When the goal is to process trace, sampling should prioritize relevance over size, representativeness, and diversity.[9] Only relevant interview subjects—for example, individuals with direct participation and/or intimate knowledge of events as per our theoretical models—can generate the high probative value data on which successful PT rests. While most qualitative scholars would probably agree that relevance matters more than size, large sample sizes are commonly reported by authors and taken by readers as signals of the quality and weight of the evidence. Moreover, recognizing that relevance matters more than size does not tell us how to determine relevance. This section tackles this issue.

We then turn to a series of recommendations about the design of interview protocols. First, process tracers should make explicit use of their theoretical models, and the fingerprints (observable implications) processes are expected to leave on the empirical record, to identify which components of the causal process interview protocols must focus on. Questionnaires should then be designed so that all those components of the theoretical process are exposed to a test against evidence. This is paramount to complying with the "completeness standard" (Waldner 2014) and corroborating the "productive continuity" of mechanisms (Beach and Pedersen 2019; Machamer, Darden, and Craver 2000; Runhardt 2016), which *some* approaches to PT see as best practice.[10] Second, structuring protocols so that the order of the questions tracks the progression of the causal chain may help increase the probative value of testimonies and open up room for serendipity and discovery. Finally, interview protocols should include questions (or batteries of questions) explicitly addressing alternative explanations. This is crucial if we want to be "equally tough" on rival explanations, prevent the interview guide from leading us toward theory confirmation (rather than theory

testing), and avoid the trap of taking the "absence of evidence" for "evidence of absence" (Bennett and Checkel 2014b).

We see this paper as a part of a new wave of work on PT. While seminal work focused on the epistemological foundations of the method and developed some helpful general guidelines and "best practices," scholars have recently turned attention to more practical issues.[11] This does not mean that foundational debates have been settled. They continue on critical issues such as how to define causal mechanisms (Jacobs 2016) and the detail to which researchers should unpack them (Beach and Pedersen 2019); how inference should be understood, and what constitutes a satisfactory explanation (Bennett, Fairfield, and Soifer 2019); and the nature of causal claims and causation (Beach 2017; Runhardt 2021), among others. In fact, scholars have developed different approaches to PT that, to varying degrees, differ in these and other fundamental issues.[12] While some of our recommendations are relevant regardless of the approach to PT one embraces, others might apply more to some approaches than others, and yet others might be in tension with specific aspects of specific approaches. When this is the case, we make it explicit. Similarly, some of our recommendations might depart from advice in the general literature on interviewing or take specific forms when applied to PT. We also make this explicit.

Even if it is widely recognized that the best methods for data analysis will not produce strong results if valid or relevant data aren't generated, relatively little has been written on data generation in the growing literature on PT. Zaks (2021: 72) rightly notes that process tracers have more access to training in data analysis than in the process of collecting that data in the first place. Our discussion addresses this imbalance and encourages process tracers to think more systematically about their interviewing plans at the design stage. This should increase the probative value of the insights they use to evaluate claims about the presence and operation of causal processes. While some readers might feel that our recommendations risk imposing too much structure on a data generation technique that is strong in part because it offers flexibility, we argue that this structure should not hamper serendipity and discovery. On the contrary, it can open room for it. Small and Calarco (2022: 1418) propose seeing the data collection process as divided into two components: conception and execution. Building on this view, we argue that good conception can boost discovery during execution. Other readers might find some of the issues discussed here obvious. To be sure, much of what we say is tacitly understood and often intuitively practiced. Articulating it explicitly, however, is important to create awareness and

stimulate conversations about best practices among seasoned practitioners and, especially, to offer guidance to those new to the method.

The recommendations we outline below are based on our experience using interviews in the context of PT studies and our reading of existing research that combines interviews and PT. In contrast to quantitative scholars, who can rely on simulations to assess the benefits of new estimators or strictly replicate published work to see how results vary with the use of alternative modeling specifications, this kind of test is not commonly available for qualitative methodological advice. Throughout this paper, we cite multiple examples from our own work and that of other scholars, covering a variety of subfields to illustrate the logic and potential advantages of the strategies we propose. This, however, does not mean further evaluation of their merits isn't necessary. Practitioners should actively probe whether implementing our recommendations improves the analytic quality of interview-based PT.

PT and Interviewing

Regardless of the approach one takes, PT rests on the idea that causal mechanisms are essential for good causal explanations (Bennett, Fairfield, and Soifer 2019). A central task in PT is, therefore, to unearth evidence of the presence and operation of mechanisms in the cases under examination. The centrality of mechanisms in PT is crystal clear in the two go-to manuals on the method. In Beach and Pedersen's (2019: 1) book, PT is defined as a method for tracing causal mechanisms using empirical analysis of how causal mechanisms operate in real-world cases. Similarly, in the Introduction to their edited volume, Bennett and Checkel (2014b: 3-4) argue that techniques falling under the label of PT are those well suited for measuring and testing hypothesized causal mechanisms. Even in PT approaches that do not consider it necessary to theorize the minutiae of causal processes and trace each component in detail to achieve "completeness" for rigorous inference, scholars still highlight that "a well-specified explanatory hypothesis should generally include some sort of causal mechanism" (Fairfield and Charman 2022: 80).[13]

When conducting PT, particularly for theory-testing and theory-revision purposes,[12] it is essential to begin crafting a theoretical account that specifies the causal pathway one believes leads to the outcome of interest. One way of thinking of these pathways is as consisting of *steps* and *mechanisms*. "Steps" refer to the theoretically relevant events that need to happen to move from the causal condition to the outcome of interest. The steps included in the theoretical model must provide a logical account of the states of affairs, decisions, or actions/reactions required to produce a transformation. For example, in a

classical model of democratic transitions, steps may include deteriorating economic conditions, a split in the ruling coalition, mass protests, etc. (O'Donnell, Schmitter, and Whitehead 1986). "Mechanisms" are what "makes the system in question tick" (Bunge 2004: 182), creating links between these steps. In the social and political world, mechanisms usually —but not always—refer to the reasons why actors with certain preferences, motivations, and capabilities behave in certain ways and react to the actions of others or to changes in things happening around them.[14]

Mechanisms are critical for PT because they help account for the progression of a case along the causal chain, specifying why and how actors transmit causal energy from one step to the other. Examples of mechanisms from PT research include Wood's (2003) "pleasure of agency," which she uses to explain peasant support for insurgents in El Salvador; Cramer's (2016) "resentment," which she uses to explain why people in rural America interpret political events in ways that lead them to develop a preference for small government; and Closa and Palestini's (2018) "tutelage," which they use to explain why decision-makers formalize enforceable democracy clauses in two Latin American integration schemes. Some scholars go even more "micro" in their quest for mechanisms. For instance, Holmes (2018) explains signal credibility during face-to-face diplomatic encounters with reference to what neuroscientists call the brain's "mirroring system."

The actors engaging in these activities do not need to be individuals; they could be interpersonal networks, social groups, or institutions. As such, mechanisms aren't exclusively cognitive—changes in preferences, perceptions, and meanings; they can also be relational and interactional, connecting people (McAdam, Tarrow, and Tilly 2001; Tilly 2001). For example, "brokerage" or joining hitherto disconnected actors through the intervention of third parties plays a central role in explaining macro-outcomes such as the transnationalization of social movements (Della Porta and Tarrow 2005).

This take on mechanisms, and the centrality they have in PT research, suggests that theoretical models in PT are commonly populated by actors who act and react based on changes in their internal cognitions, relationships with other actors, or in their environment.[15] As such, theoretical models often cry for using interviews to document at least parts of the process linking a cause (or set of causes) with the outcome of interest. Indeed, interviews are not only particularly well suited to study actors' motives, perceptions, and meaning (Small and Calarco 2022; Small and Cook 2023). They are also particularly helpful for uncovering social processes (Lareau 2021, Chapter 1), as they supply "account evidence"—recollections of what happened and why— and "sequence evidence"—the order in which events took place (Beach and

Pedersen 2019: 172). It is therefore unsurprising that Beach and Pedersen (2019: 213) report that "one of the most commonly used sources of mechanistic evidence in process tracing research is interviews."

Elite and nonelite interviews afford researchers the opportunity to interact directly with the actors that populate models, such as high-level decision-makers in Closa and Palestini (2018) or peasants in Wood (2003). These interactions grant access to actors' complex decision-making and meaning-making processes. Even when other forms of data like surveys, field notes from direct observation, or historical documents (e.g., declassified meeting minutes, biographical accounts) are available, and more central to one's research, such sources often raise new questions and leave others unanswered. Interviews are an excellent complement because they provide a window into individual thought processes that are not directly observable, or into informal backroom interactions for which written records commonly do not exist. With such a direct window into aspects of crucial importance in processes of social action and interaction, process tracers do not have to infer motives, preferences, perceptions, or meaning from observed behavior.

Some of these characteristics make interviews a good source of evidence for qualitative research in general. However, what makes the recommendations that follow specific is that, unlike interviews in general, interviews in PT are geared toward unearthing observations to document and evidence processes linking causes to outcomes, with a particular emphasis on what some PT scholars have called "causal process observations (CPOs)" (Brady 2010; Brady and Collier 2010; Collier 2011) and "mechanistic evidence" (Beach and Pedersen 2019: Chapter 5).

Sampling

Deciding "whom to talk to?" is a primary concern among those who seek to make valid descriptive or causal inferences using interview data. It is therefore not surprising that sampling techniques occupy a central place in the literature on interview methods in sociology, political science, and anthropology.[16] In this section, we discuss some of the implications that the specificities of PT might have for sampling procedures. We make two core points: (1) relevance—rather than quantity, representativeness, or diversity—is the criterion to use when selecting respondents, and (2) the probative value of empirical observations—rather than quantity—is the criterion for evaluating interview data. These two points are related: researchers are more likely to get testimonial data with high probative value from relevant respondents. To do so, we begin with a general discussion of the "relevance"

criterion and then move on to identify ways to maximize it. First, we tackle the challenges of purposive sampling for PT, discussing how to identify *who* are the relevant informants. Second, we delve into the implications of the relevance criterion for deciding *when* to interview informants. We make the case that sequencing interviews so that they match the progression of the causal chain (the steps in the process) can lead to more targeted testimonial observations and give the researcher better information to judge the quality of these testimonies. Finally, we discuss "stopping rules" and their implications for relevance.

Relevance

Critics of the interview method often argue that it has "low scientific value" and that it is "more folklore than truth" because the data interviews produce are characterized by "extreme particularity." As Kapiszewski, Maclean, and Read (2015: 200) report, critics "find it hard to imagine that theoretically interesting or generalizable claims could be constructed on the bases of the fine-grained, personalistic information collected through oral history interviewing." In light of this belief, it is way too common for scholars to highlight the *quantity* of interviews they conducted and the *size* and *representativeness* of their sample to defend the value of their testimonial data.[17] This is understandable: interviewing a very large number of subjects may make it more likely to publish in certain journals (as implied in the guidelines of *Gender & Society*), satisfy reviewers coming from a quantitative research tradition, increase chances of securing research grants, confer authoritativeness (the so-called "ethnographic authority") or signal high degrees of "exposure."[18]

Professional pragmatism aside, concerns about the "extreme particularity" of the data misunderstand the value of interviewing in general and for PT in particular. The assumption driving this concern is that a social scientist's main goal is to generalize from a sample to a population. This type of inference is indeed central in most survey research. Consequently, some influential publications implicitly and explicitly suggest that researchers should apply survey research criteria (such as sample size, selection with known probabilities, and sample bias) to interview research (King, Keohane, and Verba 1994), or should turn to purposive sampling only when lists of potential participants for random sampling procedures are not available (Knott et al. 2022). Yet, this type of inference is often *not* at the heart of most interview research (Small 2009), let alone of studies that employ PT.

The PT-specific inferential challenge is to show that the data one has collected warrants the conclusion that a particular causal process indeed

materialized, leading to the outcome of interest in a specific case. In line with scientific realism, PT assumes that while causal structures, processes, and mechanisms may be part of the unobservable ontological "deep" (Baert 1998; Bhaskar 1978), their operation leaves fingerprints in the empirical record that researchers can trace. Regardless of how much we unpack the theorized process into its constituent parts (which varies across different approaches to PT), process tracers draw observable implications from these theorized processes and seek and evaluate evidence to establish whether the process occurred or not as posited, or whether it matches better one explanation than its rivals. In interview records, relevant traces could take multiple forms, from facts the interviewee reveals about an event they were involved in, to how they discuss their experiences and perceptions, including metadata (silences, sighs, etc.).[19]

In PT, researchers make within-case inferences based on how well empirical observations match the expected fingerprints of the theory in a particular case. As such, "extremely particular" interview data may be exactly what we need in PT. In contrast to many interview research situations where the number of potential interviewees who fit the "inclusion criteria" exceeds the number one can actually interview (Knott et al. 2022), process tracers are often after "fingerprints" that only a few particular actors could reveal. This is often the case because in PT theoretical models commonly require very particular evidence, such as the motives behind known decisions by particular actors or the perceptions of experiences lived only by specific individuals. The implication is that sometimes our pool of potential interviewees is not only not representative of the larger population, but also rather small in size.[20] For example, in Closa and Palestini's (2018) work, the authors had the task of tracing a high-level decision-making process. Consequently, only a limited number of individuals were likely to be relevant. They could have interviewed an impressively large set of people and still missed the relevant actors as per their theoretical model, thus spending precious time talking to a host of irrelevant—or less relevant—people.

While this point is more straightforward in elite interviewing, it also applies to nonelite interviews. Consider, for example, Masullo's (2021) research on why some Colombian peasant communities nonviolently refuse to cooperate with armed groups during civil war. The theoretical model stressed the role of local communal leaders in shaping this decision. To probe this claim, there were two most relevant classes of actors to be interviewed: first, a limited set of people who were community leaders when the choice to engage in nonviolent collective action was made, and second, a larger group of villagers that were present when the choice was made. To

sample these two groups, random sampling or population-based sampling didn't make sense: some villagers needed to have a higher chance of being interviewed (e.g., community leaders) and many others had to have little or no chance of being selected (e.g., people who moved to the village after the decision was made). After all, Masullo's goal was not to establish average opinions in the village but to uncover a decision-making process by a delimited set of actors in a particular time and space. Moreover, as the choices he was studying took place years (in some cases, decades) ago, it made little sense to construct a sampling frame of the village at the time of research and draw a representative sample of residents. Doing so would have likely been counterproductive, as he might have missed some critical interviewees and spent considerable time speaking to less relevant ones.

In both examples, the researchers were concerned with within-case inferences and not with generalizing from the sample of interviewees to a larger population, making the representativeness of the sample irrelevant. To be sure, some process tracers might well want to probe the generalizability of the mechanisms traced in, say, one country, to a larger population of similar countries. However, sampling procedures for within-case analysis via interviews do not play a part in this generalization exercise. Instead, the plausibility of such generalization hinges on careful case selection strategies to trace processes in other cases, such that the comparison illuminates the portability of mechanisms and processes (Beach and Pedersen 2018; Lyall 2014; McAdam, Tarrow, and Tilly 2001).

Methodological texts on interviewing commonly stress the importance of interviewing a wide diversity of actors (Knott et al. 2022; Mosley 2013). While this sampling advice might make sense in general as more viewpoints are better than fewer or skewed viewpoints, when sampling for PT, relevance should come before diversity. In fact, well-specified process theories might occasionally point to limited diversity. Take, for example, González-Ocantos' (2016) research on judges' increasing reliance on international law to punish human rights criminals in Latin America. The author argues that this was the outcome of a process of legal reeducation ignited by non-governmental organizations (NGOs). Probing this argument required establishing that NGOs (i) understood judicial inaction as a problem of knowledge about international law, not just of lack of political will to punish the military, and (ii) that NGOs targeted judges with pedagogical interventions to remedy these deficits. Materials from archives of victim organizations corroborated that NGOs chose to spend time teaching judges international human rights law, but they did not say much about *why* they did so. Therefore, interviews with the actors involved in designing

these interventions were crucial to document the mechanism leading to this step in the causal process. Not all past and present members of staff could provide relevant insights, so casting a wide net to capture the nuances of the NGO world and give voice to a wide diversity of viewpoints was beside the point. Instead, the archival documents suggested a very limited pool of NGO leaders had planned and designed the reeducation campaign.

Purposive Sampling: How to Identify Relevant Interviews?

Given the general PT aim of making inferences about hypothesized causal processes in specific cases, process tracers are generally well advised to work with nonprobability/nonrandom samples of interview subjects (Tansey 2007). Sampling in PT is therefore most often *purposive*. While it is commonly assumed that with some dose of common sense researchers can develop purposive samples that suit their needs, we maintain that purposive sampling in PT is not just a matter of common sense. In PT, the theoretical models we develop have a strong say in how we *purposively* select participants.[21] As we have seen in the cited examples, theoretical models in PT are often populated by actors engaging in activities (Beach and Pedersen 2019). In addition to pushing the causal story forward, these actors are responsible for generating the data we need (Gonzalez-Ocantos and LaPorte 2021: 1410). Therefore, they are likely the most *relevant* people to talk to because they can provide us with the "account evidence" and "sequence evidence" necessary to more effectively evaluate whether the causal chain indeed materialized (Beach and Pedersen 2019: 213-15; Tansey 2007: 765). In short, the capacity to offer testimonial data with high probative value makes actors relevant, and this should be our chief selection criterion.

Well-specified theoretical models that clearly identify the actors and activities that populate causal chains offer a good indication of the most relevant classes of actors we should interview to test whether the process was present in the case and unfolded in the way we theorized it. For example, in Winward's (2021) PT study of mass killings in Indonesia in the mid-1960s, the first step in the theorized process linking low state intelligence capacity to mass categorical violence, states that security forces approach local civilian elites for the information they miss given low capacity. Here, the theory clearly points at two relevant classes of interviewees: local elites and security forces. Similarly, when working within PT approaches that emphasize the importance of contrasting alternative explanations, well-specified alternatives will also indicate relevant classes of actors we should

interview to give those accounts a real chance. This is crucial, as the actors populating our proposed model might not be the same as those engaging in the activities featured in rival explanations (we discuss alternative explanations in greater detail in a later section).

To move from models to sampling, we first need to sketch in our models the classes of actors that engage in key activities—for example, peasants living in warzones, middle-class protesters, high-level officials, etc. Then, we need to move on to "operationalize" them in terms of proper names. This involves, for example, moving from "senior decision-makers who participated in the design and adoption of democracy clauses in regional integration schemes" to "Sanguinetti, Lacalle, Valdés, Insulza in Mercosur and Unasur" (Closa and Palestini 2018). While the examples we have provided identify the classes of actors in a more or less straightforward way, theoretical models may sometimes take us away from the "usual suspects." For example, in a study exploring how middle-class children enjoy educational advantages, "common sense" would probably tell us that children and their teachers were the key relevant actors to interview. Yet, if our theoretical model posits that lobbying is a critical component of the process because it helps secure advantageous school districts for middle-class children (Lareau, Weininger, and Cox 2018), our "common sense" purposive sample would miss the most relevant actor: lobbyists. Similarly, moving from the class of actors in the theoretical model to proper names in a specific case is not always straightforward. Consider, for example, the case of Kramer's (1990) study examining the Soviet Union's decision-making process prior to the Cuban Missile Crisis. The model would likely be populated by "high-level officials" from the Soviet side directly dealing with US-related issues. As such, when moving to "proper names", Soviet ambassador to the United States during the crisis —Anatoly Dobrynin—would seem like a highly relevant actor. As the author notes, however, Dobrynin was not involved in the highest-level deliberations and learned about the deployment of the missiles only *after* the United States had discovered them. Moving from relevant classes of actors to actual interviewees often requires excellent contextual knowledge.

While relevance might vary from context to context and according to the specificities of the theorized process, three general criteria can help us identify relevant informants. First, those with the most direct involvement in the events and activities that make up our theoretical models. Second, those sufficiently close to either these actors or the process itself. This includes people who might have acquired first-hand information from the main actors in the model, as well as close observers. Third, people temporally too far removed from the process are less relevant than those who are temporally closer to the

process.[22] While talking to those directly involved in the process is often ideal, also talking to other actors who had indirect ways to access key information might be valuable. In some circumstances, these actors can be even more relevant, as those not directly involved in the process may be not only more easily accessible but also more likely to talk and to talk frankly. Direct involvement in the process might lead people to lose perspective or create personal incentives to hide or manipulate evidence. Put differently, direct involvement in the process is a strong indicator of relevance, but one the researcher needs to assess carefully both in the abstract and in relation to the specifics of the case, especially with access and reliability concerns in mind.[23]

Sequencing Interviews: When to Interview Different Classes of Actors?

Relevance as a key sampling criterion, and the practice of allowing our theoretical models to inform sampling, also provide guidelines on *when* we should interview different classes of actors. We contend that in PT there are potential advantages in sequencing interview sampling, that is, selecting relevant participants for each step of the causal chain and interviewing them following the order of these steps.[24] This is mostly relevant in PT approaches that advocate documenting and evidencing all the steps in the causal chain (or all the parts into which the process was disaggregated), including the mechanistic system approach, as well as in variants of PT that give primacy to the "completeness standard" and/or to "productive continuity" (Beach and Pedersen 2019; George and Bennett 2005: 223; González-Ocantos and LaPorte 2021; Runhardt 2015, 2016; Waldner 2014, 2015). According to some of these approaches, the inability to document all components of the process leaves important explanatory gaps or breaks the productive continuity. Missing data therefore invites the researcher back to the drawing board to rethink the causal chain (Machamer, Darden and Craver 2000: 3). Consequently, if the researcher cannot establish the first steps in the sequence, classes of interviewees previously identified as relevant sources of information about events further down the causal chain, suddenly become irrelevant or not a priority—at least until the researcher reassesses the project.

The main advantage of adopting a sequential approach is that it facilitates identifying relevant respondents and fine-tuning the quest for information as one goes along in ways that align with the requirements of (some) approaches to PT. The point is a rather simple one. A component part of the theorized

process will typically involve actors doing things to other actors, thus enabling the flow of "causal energy" and pushing the chain to the next step. If we first obtain information about the activities and actors instigating the process (e.g., Step 1), we could have a better sense of what information to look for next (e.g., Step 2), including who were the relevant targets of those activities (and therefore people we ought to interview) and what they were reacting to. Each interview focusing on one component of the sequence will provide an increasing understanding of how to approach interviewing, including who to interview and what to ask, for the subsequent components. An implication here is, therefore, that one might need to sample different actors (or classes of actors) for different components of the process, but also that different interviewees will be subject to different interview protocols. In addition, sequencing might help us learn about aspects and actors of the process that we might have missed in our initial theorizing, opening room for discovery. This might alleviate concerns that allowing the theoretical model to inform our approach to interviews could make the process too rigid, threatening some of the key advantages of interviewing such as iteration, flexibility, and serendipity.[25]

Kaplan's (2017) study of how community organizations in Colombia protect themselves from civil war violence offers an illustrative example of the advantages of sequencing while sampling and interviewing. The author was interested in understanding how armed groups *reacted* to different local organizational structures and protective strategies—a process that ultimately shaped levels of violence toward communities. The theory implied that communities first established local institutions to protect themselves and then armed groups decided to show restraint as they saw civilians' organizational capacity. Consequently, he first conducted interviews with organized civilians to understand the first step (establishment of local institutions) and, *afterwards,* he interviewed former combatants to understand reactions to it. Insights from interviews with civilians were crucial to arrive at the next set of interviewees, as they offered a clearer sense of the reality that armed groups faced on the ground. Without first knowing what civilians had done to protect themselves, Kaplan would have had a harder time knowing what protective campaigns to ask former combatants about. In addition, sequencing allowed him to evaluate potential biases in the testimonies of former combatants, which he deemed less reliable. Having first talked to civilians, he was in a good position to identify possible misconstructions or omissions in combatant testimonies, and presumably probe those biases during interviews with more targeted or incisive questioning (Kaplan 2017: 107). Had he proceeded the other way around, it is possible that he would have had to re-interview combatants for triangulation purposes. This would have

introduced inefficiencies to the data collection process, and of course, he would have risked being unable to re-contact key sources.

Defining a Stopping Rule: When do Additional Interviews Become Irrelevant?

The emphasis on relevance as a sampling criterion over quantity, representativeness or even diversity is intimately associated with the notion of probative value in PT and the method's maxim that not all pieces of evidence have the same evidentiary weight in a given context (e.g., Beach and Pedersen 2019; Bennett and Checkel 2014b; Fairfield 2015b; Mahoney 2012; Mahoney and Vanderpoel 2015). A key feature of evidence evaluation procedures in many approaches to PT (particularly in the Bayesian tradition), such as the so-called "PT tests" (straw in the wind, hoop, smoking gun, doubly decisive) and the logic of "certainty" and "uniqueness," is that the inferential power of the evidence is not judged in terms of quantity but quality.[26] To put it simply, as proponents of explicit Bayesian PT stress, "evidence is to be weighed, not counted" (Fairfield and Charman 2022: 160). If we formulate theoretically unique and unlikely expectations about fingerprints that we can find when interviewing representatives of a class of actors, we might collect a few highly probative pieces of testimonial evidence such that we might not need to continue looking for additional evidence. We can stop interviewing. If a plane suffers an accident, and the surviving pilot tells the investigators that he deliberately tried to crash the plane, a detective may not need to look any further to draw inferences about who was materially responsible for the accident. It is unlikely that the pilot is not telling the truth—not only because of his privileged role in the process, but also the high personal costs of recognizing responsibility.[27] Additional interviews with passengers, air controllers, etc., could be reasonably deemed less critical.[28]

Fairfield's (2013: 47-59, 56; 2015a) research on why unequal democracies sometimes tax economic elites offers a good example. In her case study of a tax reform that in 2005 revoked a regressive tax subsidy in Chile, a testimony from one relevant actor lent strong support to her hypothesis (relative to rivals) that an equity appeal by the governing center-left coalition during the presidential race secured opposition support for the reform. In an interview, a right-wing deputy explicitly noted that despite their capacity to block the reform in Congress, the opposition reluctantly accepted it to avoid harming its presidential candidates in a race where voters strongly valued equity. Statements from this interviewee not only supported the

argument in a very explicit way, but given political incentives, it also consti-
tuted evidence that was highly unlikely to exist if the incumbent's appeals had
played no role. While interviewing other deputies from the opposition coali-
tion would be useful for corroboration, many more interviews with this class
of respondents would unlikely offer much additional evidence to update
priors.[29]

In the aforementioned study of human rights prosecutions by
González-Ocantos (2016), interviews with judges were crucial to establish
the effectiveness of informal reeducation efforts. The risk, of course, was
that as members of the elite, judges would not feel comfortable admitting
prior ignorance of international law or conceding that NGOs that were also
plaintiffs in the cases had deliberately shaped their legal views
(González-Ocantos 2016: 25). Fortunately, many of the judges revealed
their prior ignorance of international law and their informal (and likely ques-
tionable) relationship with NGOs with unexpected candidness
(González-Ocantos 2016: 181). This increased the probative value of their
testimonies and made it less urgent to move on to interview court clerks,
who could in principle speak more freely but were more removed from the
activities populating the process. In other words, as the more relevant
actors—those in a position to supply more valuable insights—did offer state-
ments in line with the expected observables, this minimized the need to
expand the sample to less relevant actors to look for those observables or
to verify the statements made by judges.

Benchmarks based on the probative value of evidence rather than quantity
or representativeness offer a PT-aligned indication of when to stop interview-
ing some classes of actors or collecting interview data in general. [30] To be
sure, this type of high probative evidence is not always available, so the stop-
ping rule may be hard to apply. This is because not all observations map
clearly onto the different PT tests or fall neatly into the categories of certainty
and uniqueness (Fairfield 2015b: 48). An alternative and less demanding
PT-aligned "stopping rule" builds on what the general literature on interview-
ing calls "saturation": "You stop when you encounter diminishing returns,
when the information you obtain is redundant or peripheral, when what
you learn that is new adds too little to what you already know" (Weiss
1995: 21).[31] Saturation is relevant in PT research, especially when one has
several equally relevant interviewees within the same class of actors in a
sample. Instead of interviewing all of them, one can stop when a subsequent
interview yields redundant information. Bennett and Checkel (2014b: 27)
advise to "stop when repetition occurs." Specified in more explicit PT updat-
ing terms, stopping when interviewees belonging to the same theoretically

relevant class offer similar testimonies makes sense because new, similar testimonies will have lower probative value. Again, corroboration needs aside, more testimonies (quantity) from the same class of interviewees do not imply stronger overall evidentiary support.

<p style="text-align:center">***</p>

In sum, interview sampling should prioritize relevance, not quantity, representativeness, or diversity, as this should help maximize the evidentiary weight of the data we collect and live up to the probative standards of PT. This, however, does not mean that sampling techniques that allow researchers to draw diverse, large, and/or representative samples from a class of actors are never called for. First, relevance is not always obvious. Sometimes we stumble upon vital actors we did not imagine existed or mattered for the outcome. Being receptive to these surprises (and following up on them) is a crucial feature of good fieldwork (Small and Calarco, Chapter 4). Similarly, there are cases where it is hard to operationalize classes of actors precisely. This is particularly likely in processes involving nonelites and when working with collective actors, such as insurgent movements, resisting communities, undocumented migrants, or the urban poor. In these cases, important decisions or actions that contribute to the outcome are not necessarily the purview of a limited pool of participants with clearly defined institutional roles (Fujii 2017: 38). When sampling for projects including these actors, there may be some "safety in numbers."

Second, certain components in the causal chain could require documenting central tendencies or shifts in average opinions. In those cases, striving for some degree of representativeness makes sense. If sampling techniques to achieve representativeness, such as those used in survey research,[32] are not practically possible or inadequate, process tracers may use other techniques of purposive sampling that allow them to have a pool of interview subjects that capture a diversity of voices and viewpoints. For example, sampling for range, where the researcher identifies subcategories of the group under examination (say residents of a refugee camp), and makes sure to interview a given number of people in each subcategory (Chan Tack and Small 2017; Small 2009; Weiss 1995). Campbell's (2018) study of the success of peacebuilding operations illustrates one of these options. She argued that alliances between international organizations' country-based staff and local stakeholders were essential for success. She needed a relatively large and diverse sample to empirically evaluate this claim. Consequently, she used a stratified purposive sampling strategy to make sure that she was gathering data from different subsets of interviewees: "staff at all levels of seniority, key former

staff from the case study organization, headquarters-based staff, and local partners of each case study organization in Burundi" (Campbell 2018: 28-29).

Designing Interview Protocols

The theoretical models that we build for PT have a bearing on how we design interview protocols. In this section, we discuss how these models (i) help maximize *coverage*—that is, identify the components of the causal process interview evidence can help evaluate and, therefore, what interview protocols should prioritize in terms of questions/batteries of questions; (ii) call for decisions about question *sequencing*—that is, offer guidance regarding the order in which questions could be more effectively asked to boost the probative value of testimonies; and (iii) demand *openness* to disconfirmation—that is, calls for the inclusion of questions on alternative explanations.

Coverage: Identifying Interview Suitability and What to Ask

In PT, the researcher first sketches a theoretical process that is believed to link cause and outcome. This can be done in a minimalist way, specifying a pathway as a whole, or in a more disaggregated form, specifying different constituent components of the pathway.[33] Then, the researcher brings contextual specificity to document a sequence of concrete events in the case under examination that are compatible with the components of the sketched theoretical process. From here, she moves on to identify the observable implications—the fingerprints—that the different components of the process might have left in the empirical record. This procedure leaves us in a good place to identify which parts of the process interview data can effectively help evaluate.

First, the *type of mechanisms* that populate our theoretical models gives us a good indication, as not all types of mechanisms cry for interview data. When mechanisms refer to individual or collective actors, including their internal cognitions, behaviors, and how they interact with other actors, interviews will likely be an appropriate data source. When they refer more to physical and social structures or environmental conditions, interviews might be less relevant, unless the mechanism relates to how these social structures constrain and facilitate the actions of relevant actors, or how cultural settings shape their cognitions.

Scholars have proposed different taxonomies of mechanisms.[34] One that has been widely used across political science and sociology is that proposed by Tilly and collaborators (2001, 2001). They identify three broad types of

mechanisms: cognitive, relational, and environmental. Traces of *cognitive* mechanisms, which include actors' normative commitments, emotions, preferences, or incentives, might require us to conduct interviews. They are not only virtually unobservable and unlikely to be recorded in writing, but they speak to aspects that interviewees are in a privileged position to self-report. Something similar can be said about *relational* mechanisms, which speak to connections among people, groups, and interpersonal networks and include activities such as persuasion, lobbying, coalition formation, or brokerage. Evidence for or against their presence and operation is also commonly identifiable in testimonial data. However, this is not necessarily the case for *environmental* mechanisms like resource depletion. As they refer to externally generated influences on conditions affecting social life and apply not to actors but to their settings, other sources of data might prove more appropriate to evaluate their presence and operation.

Second, the *observable implications* of our theorized process also offer a good indication of the components interview protocols should focus on.[35] To be sure, most of the literature on interviewing stresses the importance of structuring interview questionnaires in a way that helps us focus on *relevant* aspects. For example, Lareau (2021: 63, 74, 94) notes that "you want the interview to stay focused on what is important to you" and stresses that one's research question and research objectives provide guidance on what is most important to ask. This general advice of course applies when interviewing for PT. However, what is different is that *focus* is dictated by a well-specified theoretical model about a process; in particular, by the fingerprints that the researcher theorizes that the operation of the process should have left on the empirical record.[36] The assumptions we make about these potential fingerprints offer crucial information to determine which types of data (interview, archival, etc.) are appropriate sources of evidence. Can a respondent articulate this observable manifestation in an interview? Is this something people can self-report on? If they do, would this constitute reliable data? In short, it is fair to assume or expect that fingerprints that refer to actors' motivations for action, their perceptions and understanding of themselves and their circumstances, and the meaning they attach to what they and others do, can be found in testimonial data generated via interviews (Small and Calarco 2022; Small and Cook 2023).

Finally, the theoretical model will give you a sense not only of what to cover via interviews but also of what is a priority. Conducting interviews implies managing multiple issues simultaneously and making important in-the-moment decisions (e.g., "Should I probe this or that further?" "Should I skip this question?"), which most likely will have an impact on

the data generated (Small and Calarco 2022: 14; Lareau 2021). To make the practice less nerve-wracking, the general literature on interviewing stresses the importance of having clarity about your highest priorities (what you want to know and why) and of thinking in advance about what is "nonnegotiable" in an interview (Lareau 2021: 97-101). Many of these decisions can't be fully anticipated even when working with a well-specified process theory. Yet, while we will still have to do a lot of thinking on our feet, allowing the theoretical model to guide what components of the process the interview should focus on and what questions you must ask a given interviewee, will make for a more productive experience. Lareau (2021: 101) advises interviewers to have a "system" that helps them remember to ask the priority questions. When interviewing for PT, a clearly sketched theoretical causal process, with a clear understanding of its various fingerprints, provides us with that system.[37]

Löblová's (2018) study of the process by which epistemic communities influence policymaking provides a good illustration of how a well-sketched process theory offers indications of what aspects should be covered with interviews. Her hypothesized causal chain goes from the formation of an epistemic community to the adoption by decision-makers of a policy that is in line with the preferences of that community, with a series of intermediate components: policy promotion, consolidation of bureaucratic power, and persuasion of decision-makers (Löblová 2018: 165). Sketching these different components and deriving observable implications indicated her that interviews, sometimes in combination with written sources, were apt to look for evidence for some but not all components. In particular, the fingerprints of the third component—the consolidation of bureaucratic power by the epistemic community—called for interviews to probe whether epistemic communities indeed managed to get the attention of decision-makers. In fact, interviewing both members of epistemic communities and decision-makers revealed evidence of informal attempts to "get the ear" of decision-makers (e.g., dinner parties, informal meetings). This component of the process was relational in nature, and evidence for it was not reported in other nontestimonial sources she analyzed, such as the resumes of epistemic community members.

Similarly, in Masullo's (2017) research on wartime civilian collective action, the theoretical model sketched a process in which (a) changes in territorial control led to (b) increases in the levels of violence against civilians which, in turn, (c) altered the way that civilians perceived such violence. Perceptions of violence being unavoidable, ultimately shaped communities' choice to mobilize in opposition to armed groups despite the risks involved

in this course of action. To establish whether there was a shift in who controlled a given territory and whether increases in violence followed, the author used time-series data clashes between armed factions, homicides, and massacres. While this provided evidence for the first two components, it did not say much about how civilians *perceived* these changing war dynamics and how those perceptions shaped their choice to oppose armed groups. For these last components, the researcher used interviews, and the protocols prioritized questions about cognitive and relational mechanisms, including changes in perceptions about violence, beliefs about armed groups, and risk assessments.

Before moving to the next point (sequencing), it is important to stress that allowing your theoretical model to inform the questions you ask is not equivalent to structuring protocols such that you maximize the chances of finding *supporting* or *confirmatory* evidence. Quite the contrary: in addition to adding questions addressing competing explanations (see below), maximizing the number of theoretical parts to be assessed against interview evidence makes it harder for the researcher to establish that the causal process occurred as posited by the theoretical model, even when "chasing" confirmatory evidence. As Waldner (2015) notes, specifying the causal chain not only offers more fine-grained knowledge of the proposed theory (and, eventually, of the case) but also yields additional opportunities (and requirements) for theory falsification.[38]

This implies that PT is highly demanding in terms of coverage. This is particularly the case in PT approaches that emphasize the importance of empirically tracing the entire process for making inferences. In a foundational piece, George and Bennett (2005: 207) stressed that "*all* the intervening steps in a case must be as predicted by a hypothesis […] or else that hypothesis must be amended" (emphasis in original). Similarly, Beach and Pedersen (2013: 69) write that PT is successful when the evidence allows us to "infer that all of the parts of a hypothesized causal mechanism were actually present in a case." This coverage requirement is echoed—and even pushed further—by PT scholars who emphasize "completeness" (Waldner 2014, 2015) and "productive continuity" (Runhardt 2016). Interview protocols should, therefore, include questions to test not only whether each step and mechanism in the theoretical process took place as theorized, but also whether each step in the process is casually *linked* to those that precede and succeed it.

To illustrate this point, consider Bakke's (2013) research on the radicalization of tactics during the Chechen wars. Her theoretical argument posits that via relational and mediated diffusion, transnational insurgents influenced local Chechen fighters to use more radical tactics outside of their original

repertoire, like suicide attacks. She offers convincing evidence of a change in local fighters' acceptance of direct violence against civilians over time, shows that transnational fighters set up training camps *before* the tactical change, and that these camps were used to promote and teach new and more radical tactics. Her evidence supports that the different steps in the process did take place and that the chronology of the process indeed materialized. Despite marshaling an impressive amount of evidence, to corroborate that the steps in the process were causally linked, we would need evidence that local insurgents were convinced to use more radical tactics in those camps and thanks to teaching by transnational insurgents. Interviews with local insurgents, with questions on whether they attended these camps, the content of the teachings offered there, and their perceptions of those teachings, could be particularly helpful to unearth evidence of the (causal) links between the component parts of the theorized process. [39]

Sequencing: Deciding How to Structure Interview Protocols

Tracing processes involves uncovering evidence about how the process unfolded, that is, observations about the order in which events occurred. This is the case because many process theories make temporal claims and predict specific sequences of events, stressing that the spatiotemporal organization of actors and the activities they perform in a given context matter.[40] Temporal information contained in the evidence (Fairfield and Charman 2022, Chapter 1), or what Beach and Pedersen (2019: 171–72) call "sequence evidence," is often important for inference in PT. [41]

An implication for interview research is that organizing protocols so that they follow a temporal chronology or sequential ordering can prove helpful. This is a common technique in life history interviews and narrative interviewing, where researchers ask people to discuss events and experiences as they took place over a temporal arc from past to present (DeLuca, Clampet-Lundquist, and Edin 2016; Gerson and Damaske 2020: 82–94). However, what determines the "temporal arc" in PT is not an individual's life course, but the theoretical model specifying how the causal process supposedly unfolded. For example, an implication of Bakke's argument cited above is that local insurgents changed their tactics *after* their transnational counterparts established the camps and taught them new things. Similarly, an implication in Masullo's theory of civilian opposition to armed groups is that an increase in violence against civilians took place *before* communities decided to mobilize. If we use interviews to collect testimonial data to

evaluate these claims, evidence should ideally speak not only to the presence of each step but also to the chronology of events.[42]

Ordering the questions in the protocol so that they follow the trajectory of the model is also helpful for enhancing the probative value of testimonies that reconstruct specific components within the sequence. Sequencing questions provides interviewees with a framework for discussing not only events and activities, but also the relationship between them. It allows them to reconstruct the causal process without being prompted to do so, which, if it happens, enhances the value of their testimony. If we start with a question that asks about the first step in the causal chain and respondents themselves make a connection between this step and the next, we can be more confident in that part of our theory than if we explicitly ask them about the presence of a subsequent step and its connection to earlier ones.

To start thinking about how to formulate these questions, general advice from the literature on in-depth and ethnographic interviewing is helpful. Scholars from these traditions recommend starting interviews with "grand tour questions" (Lareau 2021; Leech 2002; Spradley 2016). These are general questions phrased in a nondirective manner to allow respondents to talk freely and encourage them to ramble on (e.g., "Could you walk me through a normal day in the village?"). These questions help ease the respondent into the conversation and set the tone of the interview. In PT, "grand tour" questions can also be a good way to start interviews, yet these questions should be more specific and try to already direct the conversation toward describing the process we want to reconstruct.[43] While still striving to make our respondents comfortable, the primary objective is to allow them to reconstruct a sequence unprompted. Consequently, unlike in general interviewing where it might be desirable to alert our subjects to what we want to know next (Gerson and Damaske 2020, Chapter 4; Weiss 1995), for these questions to be effective in PT, we should avoid giving away too much.

One way of implementing a PT-aligned "grand tour" strategy, is to begin with a question that evokes a significant event, preferably one that corresponds to the early stages of the theorized process. This can be particularly helpful when tracing processes that unfolded years before the interview, as it helps trigger memories about the onset and address recall error problems, which are common in interviews, especially when exploring motivated action (Small and Cook 2023; Wood 2003, Chapter 2). A second option is to briefly describe the triggering cause in our model and ask the respondent to describe what happened next. Finally, asking the respondent to freely offer an account of how our outcome of interest came about could also be effective. [44]

The following example from an interview with a judge who played a crucial role in the judicialization of human rights violations in his country illustrates the benefits of leveraging question ordering to collect sequence evidence with high probative value. In this case, González-Ocantos (2016: 98–99) sought to probe whether certain innovative legal doctrines advanced by the court to favor victims of repression were the product of NGOs' efforts to teach judges international law. The protocol started with a general question (Question 1) about how the judge thought about human rights cases in the 1990s, a time when judicialization had made little progress. The goal was to see whether the judge's instinctive answer was to refer to the lack of viable legal arguments as one of the key obstacles to overcoming impunity.[45] Unprompted references to this kind of obstacle would be deemed highly supportive of the argument, whereas an exclusive emphasis on nonjuridical factors (e.g., the political situation) would lend more support to alternative explanations. The remaining questions progressively moved from obstacles to solutions, in line with the order of the causal chain. With Question 2, the researcher sought to prompt a discussion of the role of orthodox legal ideas and lack of knowledge of international law as obstacles to progress should Question 1 not elicit an unprompted discussion of this issue.[46] Question 3, in turn, asked how the judge became convinced that there were viable legal alternatives to the status quo,[47] and Question 4 asked specifically about NGO reeducation efforts.[48] The importance of these efforts in the judge's "conversion" would be deemed greater if the answer to Question 3 already covered the role of NGOs' pedagogical interventions.

In this case, sequencing led to high-quality insights. The answer to Question 1 was extremely revealing of the entire process. The judge not only made references to the lack of viable legal arguments to advance the human rights cause at the start of the process but went on to highlight the crucial role played by human rights NGOs in supplying those arguments via informal interactions with the court. In other words, the interviewee told a story compatible with the theoretical argument completely unprompted. It was therefore not necessary to ask more questions. While prompted answers do not fully undermine the probatory weight of the evidence, they do suggest the need to reassess the relative importance of the hypothesized causal process in bringing about the outcome.

The final advantage of structuring questions following the sequence of the proposed theorized process is that it could help prevent us from sticking too rigidly to our explanation. Starting interviews by inviting the interviewee to talk freely about a process should open (rather than close) room for serendipity and discovery. By asking questions about the initial steps in the process

and that do not alert our respondents to what we want to know next (or what we think the next steps or links in the process are), we give room for them to take us in many, perhaps disconfirmatory, directions. This can be invaluable as it can offer hints of alternative sequences, or steps in the sequence that we did not consider when developing our theory. Even if well-specified process theories are guiding our interview protocols, we should be attentive to these hints and inductively follow up on them by adding new questions to address the new discovery in new interviews with the same informant and/or in those with other informants.[49]

Openness: Giving Alternative Explanations a Real Chance

The critical examination and adjudication of competing explanations is one of PT's foremost priorities (Bennett 2010; Bennett and Checkel 2014b; Bennett, Charman, and Fairfield 2022; Brady, Collier, and Seawright 2006; Collier 2011; Fairfield and Charman 2017; George and Bennett 2005; Hall 2008; Rohlfing 2014; Zaks 2017).[50] Doing so is vital to assess causal claims and make valid inferences thoroughly. Moreover, by preventing researchers from restricting attention to whether evidence is consistent with a single explanation, considering alternative explanations helps guard against confirmation bias. Consequently, as part of PT's "best practices," Bennett and Checkel (2014b: 18) recommend process tracers to "cast the net widely for alternative explanations."

This emphasis on alternative accounts has important implications for how we structure our interview questionnaires. Interview protocols should include questions addressing both the proposed theoretical account and alternative explanations.[51] In other words, we must collect data to link evidence to the observable manifestations predicted by the working explanation and its alternative(s). Yet, given differences in how scholars understand the nature of alternative explanations or the proper ways to articulate them, how we include (and treat) these questions might vary across PT approaches.

In PT approaches that do not require working with mutually exclusive explanations, or articulating them as such, evidence in favor of one explanation does not automatically invalidate or undermine the alternatives.[52] Consequently, researchers must search separately for counterevidence specific to rival explanations. This implies that interview questions explicitly addressing one explanation can yield testimonial data that bears on its plausibility without having implications for alternative explanations and their relative merits (Ricks and Liu 2018; Rohlfing 2014; Zaks 2017, 2021). Central to the task of giving alternative explanations a real chance is therefore asking

independent interview questions that explicitly address them.[53] To do so, we must think carefully about the type of evidence required to corroborate or invalidate each competing explanation and craft interview questions that explicitly address them. When working with these PT approaches, questions that could elicit different and additional pieces of evidence for corroboration and invalidation are required.

How exactly to include these questions in an interview protocol can vary. For example, Ricks and Liu (2018) recommend that researchers first focus on finding evidence for their working explanation and then proceed to find counterevidence for rivals. This would imply having first batteries of questions focusing on the working hypothesis and then adding more questions (or batteries of questions) addressing alternatives. Yet, one could also combine questions addressing the working hypotheses and alternatives within topical batteries of questions. The crucial requirement is to include questions that explicitly address alternatives to find separate but disconfirming evidence. Put differently, protocol design must embrace the motto "absence of evidence is not evidence of absence" (Sober 2009). If the testimonies that we collect do not yield any evidence in favor of an alternative account, we should not take it as evidence that the alternative has no explanatory value. Only when we include specific questions that can generate data relevant to the rival account, can the "absence of evidence" be reasonably interpreted as "evidence of absence."

Early work on Zones of Peace (Mitchell 2007) suggested that the presence of a triggering violent event that shocked communities was crucial in leading communities living in warzones to mobilize in opposition to armed groups. The theoretical model proposed by Masullo (2017, 2021) to explain civilian mobilization in opposition to armed groups in rural Colombia offered an alternative account that did not feature a triggering event. In an initial round of fieldwork, the researcher reports that interviews with most villagers—leaders and nonleaders alike—did not make reference to a clearly identifiable event, and the testimonies of those who did single out a specific event did not converge on the same event. Testimonies revealed that communities had suffered a great deal of violence, but there was no evidence of one particular trigger, such as a massacre or the kidnaping of a high-profile leader. The researcher could have taken this "absence of evidence" to reject the role of a triggering event. However, the alternative account and his proposed explanation were not mutually exclusive (Zaks 2017), and the interview protocol for the first wave of fieldwork did not include any questions probing the role of triggering events. Consequently, in subsequent rounds of fieldwork, he included a separate battery of questions to tap into this

alternative more explicitly and independently. This ultimately led to testimonial evidence that cast doubt on the "triggering event" hypothesis. Even when asked to think about the potential role of different triggering events, respondents who were present when the decision to mobilize opposition was made, didn't see it as a part of their decision-making process.

The situation is somehow different when working with PT approaches that require articulating rival explanations as mutually exclusive, such as explicit Bayesian PT. Here, rival hypotheses must always be compared, and evidentiary support is always relative to a specified pair of rival hypotheses (Bennett 2014; Bennett, Charman, and Fairfield 2022; Fairfield and Charman 2017). The crucial question is not whether a piece of interview evidence supports a singular hypothesis but whether it fits better with one hypothesis relative to its rivals. This is precisely what the Bayes' rule tells us: we gain more confidence in our working explanation to the extent that it makes pieces of interview evidence more plausible than a specified alternative. In other words, the researcher won't take any interview evidence that is consistent with an explanation as supporting that explanation without asking whether that evidence might be more expected under a rival explanation.

In terms of interview questionnaires, this implies that the evidence substantiating our working explanation does not come from an independent set of questions or battery of questions. All hypotheses are compared in light of *all* salient evidence (Fairfield and Charman 2017, 2022, Chapter 7). Testimonial data from interview questions that address the working hypothesis, as much as those that address its rivals, have a bearing on both hypotheses. Therefore, observations from different questions should not be treated as different subsets of evidence to evaluate different hypotheses independently. As we always compare hypotheses and evaluate likelihood *ratios*, interview questions that directly and explicitly address the world of one hypothesis will necessarily have a bearing on the alternatives.

This does not mean, however, that we should not ask questions that address the different explanations we are working with. To do so, we must "mentally inhabit the world" of each explanation and ask how expected the evidence stemming from these questions would be in that world (Bennett, Charman, and Fairfield 2022: 300–301; Fairfield and Charman 2022, Chapter 1). Hypotheses should be articulated with enough specificity not only with an eye on what interview observations would be more or less expected under each hypothesis but also to craft questions that address both "worlds."[54] Imagining hypothetical worlds in which each hypothesis is true and thinking about the fingerprints it might have left in that world, is vital to crafting the right interview questions to give rival explanations a

real chance of fitting the evidence better than the working hypothesis. While every interview observation will lead to updating, we should still seek to collect data for which our explanations make divergent predictions. Some of the interview questions pertaining to the "world" of alternatives might not be directly (or obviously) "relevant" to the working hypothesis. One could even expect that they would elicit uninformative data regarding that hypothesis. Yet, they might be crucial for assessing which hypothesis is the frontrunner. No matter how unrelated interview questions might appear to the working hypothesis, they can still favor it if the evidence elicited is more likely in that world than in that of the alternative.

Conclusion

Employing PT has implications for the way we plan and execute interview research. In this paper, we discussed some of these implications, focusing on sampling and the design of interview protocols. Our core message is that the theoretical model of the process to be traced offers helpful guidance in crucial design decisions. It is important to keep in mind that the models that guide PT are not fixed and that letting them inform our choices does not imply embracing crisp distinctions between exploration/confirmation, theory-building/theory-testing, and induction/deduction. PT is an iterative method, even when explicitly aimed at theory testing. As Fairfield and Charman (2019: 153) put it, "prior knowledge informs hypotheses and data-gathering strategies, evidence inspires new or refined hypotheses along the way, and there is continual feedback between theory and data." The researcher, therefore, commonly goes back and forth between data and theory. As new evidence enters into our analysis and our models change, we will likely need to incorporate new actors in our samples and new questions in our protocols. While offering structure, our recommendations should not prevent researchers from engaging in this type of discovery, revision, or iteration. Quite the contrary: by disciplining our data collection in the way we propose, we can structure serendipity. Put simply, we become more alert to incoming evidence that does not support our initial expectations and to the need to discard or refine theory.

Table 1 summarizes our recommendations. Regarding sampling, we emphasized the importance of prioritizing relevance over other considerations, relying on the theoretical model as a guide for sampling relevant sources.[55] We further discussed how purposive sampling, sequential sampling, and PT-aligned stopping rules can help identify and maximize relevance. In terms of the design of interview protocols, we discussed how

Table I. Summary of Recommendations for a PT-Aligned use of Interviewing Methods.

	Recommendation	Note
Sampling		
Maximize relevance, that is, the selection of individuals with direct participation/intimate knowledge	*Purposive sampling*: Well-specified theoretical models that identify actors/activities in causal chains offer good indication of relevant sources. First, those with the most direct involvement in the events/activities. Second, those sufficiently close to either these actors or to the process. Third, people temporally close to the process	To give alternative explanations a chance, respondents should be selected in light of rival accounts. Relevant actors may vary across explanations.
	Sequencing: Select relevant sources for each step of the causal chain and interview them following the sequence. Obtaining information about activities/entities instigating the process indicates what information to look for next (relevant targets of initial activities, what they reacted to). If we cannot establish initial steps, interview subjects previously identified as relevant for later steps, could become less relevant.	This recommendation mostly applies to PT approaches that promote the "completeness standard" and "productive continuity"
	Stopping Rule: Formulate specific and unlikely observable implications before	

(continued)

Table 1. (continued)

	Recommendation	Note
	interviewing. If compatible evidence is found, the researcher might not need to continue looking for additional evidence. They can stop interviewing. High probative value evidence is not always available. A less demanding stopping rule builds on the notion of "saturation."	
Designing interview protocols		
Identify interview suitability for different aspects of the theorized process	Maximizing the components of the theory assessed against evidence increases disconfirmation chances. First, not all mechanisms cry for interviews. When mechanisms refer to individual or collective actors, including cognition and behavior, but also how they interact with others, interviews will be an appropriate source. Second, the *observable implications* of the model offer good guidance. Can a respondent articulate this observable implication in an interview? Is this something people can self-report on? Would this constitute reliable data?	Maximizing coverage is particularly relevant—and demanding—for PT approaches that promote the "completeness standard" and "productive continuity."
Sequence interview questions	Question protocols should track the	

Table 1. (continued)

	Recommendation	Note
	progression of the causal chain. Sequencing helps verify the chronology of events. It can also boost the probative value of testimonies that reconstruct the sequence, especially unprompted ones. "Grand tour," questions help achieve these goals.	
Be open to disconfirmation	Interview protocols should include questions addressing both the proposed theoretical account and alternative explanations. The researcher must collect data to link evidence to the observable implications of the working explanation and its alternative(s).	In PT approaches that do not require positing mutually exclusive explanations, evidence in favor of one explanation does not automatically invalidate/undermine alternatives. Researchers must search separately for evidence specific to rival accounts. When working with PT approaches that require articulating mutually exclusive rivals, especially explicit Bayesian PT, evidentiary value is relative to this set of hypotheses. The evidence that substantiates our working explanation does not come from independent observations. All hypotheses are compared against *all* evidence, assessing which one fits best.

theoretical models can help maximize the coverage of our interviews; why paying attention to the order of questions and the use of "grand tour" questions can help boost the probative value of the data; and the importance of openness to disconfirmation via careful consideration of alternative explanations in our interview questionnaires. While a key pillar of our argument is that interviewing (as a data-collection technique) can be used regardless of one's take on PT, our various recommendations do take into account differences between the "schools of thought" that animate the methodological literature on PT. Some are more aligned with specific approaches, and others can be adapted.

In terms of future research, methodologists could push discussions about the implications of PT for data collection in various productive directions. First, there is a dearth of work exploring how one ought to report interview data to increase evidentiary value and analytic transparency. While active citation techniques (Moravcsik 2010) and the annotation for transparent inquiry (ATI) initiative offer excellent starting points, more can be said about this crucial aspect of the research process. For example, Small and Calarco's (2022) excellent book on ethnographic and interviewing methods puts forward five criteria for reporting and evaluating qualitative research. Process tracers should take their arguments very seriously and think about how their recommendations should be adapted to PT.[56] Second, interviewing is obviously not the only qualitative data-collection technique used to unearth evidence in PT research. Researchers also rely on archival methods, focus groups, and participant observation, among others. A promising avenue for further research is to explore the implications of PT for each one of these data-collection techniques. Finally, as noted in the introduction, we see our contribution as part of a new wave of methodological papers focused on practical, "how to" questions. Having said this, anyone paying attention to recent developments in the literature on PT will know that foundational debates about what PT is and how it should be conducted are far from settled. As our discussion makes it clear, a PT-aligned use of interviewing methods has some general implications but also approach-specific ones. More work should be done to map and systematize differences between, say, explicit Bayesian PT and approaches that emphasize "completeness" or "productive continuity" to identify common ground and fine-tune practical advice even further.

Author's Note

Declaration of Conflicting Interests

Funding

ORCID iD

Ezequiel Gonzalez-Ocantos https://orcid.org/0000-0002-9617-478X

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Notes

1. We would like to thank Ivan Bakalov, Imke Harbers, Diana Kapiszewski, Sidney Tarrow, Elisabeth Wood, participants at APSA 2019 and the 2022 ERC Workshop on Mixed Methods at Cambridge University, and SMR reviewers and editor for helpful comments on previous drafts of this paper. Lara Hankeln provided superb research assistance.
2. See Beach and Pedersen (2013, 2019); Bennett and Checkel (2014a); Collier (2011); Fairfield (2015b); Fairfield and Charman (2017); Goertz and Mahoney (2012); Humphreys and Jacobs (2015); Zaks (2017).
3. This is the case in George and Bennett's (2005) influential book, where the use of interviews for PT is only recognized in passing. For a discussion of this limitation, see Odell (2006) and Tansey (2007). For illustrations of PT research drawing on a wealth of primary archival sources and historical evidence, see Handlin (2017), Schultz (2001), and Tannenwald (2007).
4. We coded papers published between 2013 and 2023 in five top journals in Political Science and Sociology (American Political Science Review, Comparative Politics, World Politics, American Sociological Review, and Social Forces). Out of 183 papers that report conducting PT, 114 use interviews—that's over 60%. Results are as follows: American Political Science Review 18 out of 41; Comparative Politics 48 out of 76; World Politics 26 out of 36; American Sociological Review 15 out of 22; Social Forces 7 out of 8.
5. There is ample literature on the limitations and challenges of interview data. In a recent paper focusing on interviewing to study motivated action, Small and Cook (2023) identify five main challenges researchers face – deception, recall error,

reasonableness, intentionality bias, and single-motive bias – along with several strategies to address them. Fairfield and Charman (2022, Chapter 3) offer recommendations for how to deal with biases and other shortcomings of testimonial evidence within a Bayesian framework. In an exemplar application of PT, Winward (2021) includes an appendix discussing the degree to which the sources he used, including interviews, can be trusted.

6. Other excellent examples of PT work using interview data include: Fairfield (2013, 2015a); Fairfield and Garay (2017); Fu (2017); Holland (2017); Jacobs (2011); Lake (2018); Löblová (2018); Schimmelfennig (2003); Winward (2021); Wood (2000, 2001).

7. This literature is vast, covering different types of interviews and cutting across disciplinary and epistemological boundaries. Some influential works in Political Science and Sociology include: Berry (2002); Foddy (1993); Fujii (2017); Kapiszewski, Maclean, and Read (2015, Chapter 6); Mosley (2013); Spradley (2016), Tansey (2007) and Weiss (1995). More recent contributions, developed mostly in Sociology, include Gerson and Damaske (2020), Lareau (2021), Small and Cook (2023), Small and Calarco (2022), and Tavory (2020).

8. The two go-to manuals on PT only marginally discuss interviewing. In Bennett and Checkel's (2014a) volume, interviews are briefly discussed as a "proxy" to participant observation (Pouliot 2014: 247-48) and in discussions about transparency (Dunning 2014: 232-33). Beach and Pedersen explicitly discuss interviewing, but they do so only very briefly (134-36 in the 2013 edition and 213-15 in the 2019 edition). To our knowledge, the main exception is Tansey's (2007) short essay on elite interviewing. While we build on his work and further develop his argument for nonprobability sampling, we go beyond sampling issues, don't limit our discussion to elite interviewing, and engage with an important body of PT literature published after his seminal paper.

9. This point is extendable to interview research in general (see Small 2009; Small and Calarco 2022), but as we show in this paper, it takes specific forms in PT research.

10. In our understanding of the "completeness standard," completeness does not imply that the researcher must reach total certainty that each component of the causal chain is present in the case—but see Bennett (2016) and Bennett and Checkel (2014a, 265). As in Bayesian PT approaches, evidence can be partial and incomplete, and how strongly it lends support can vary from component to component. What matters for our discussion is that completeness requires us to subject each component of the process to a test against evidence. Different PT approaches disagree on the type of evidence (e.g., counterfactual or CPOs) that can sufficiently corroborate the component parts of the process and some stress differences between "completeness" and "productive continuity" (Beach 2017;

Beach and Pedersen 2019, 64-69; Mahoney and Barrenechea 2019; Runhardt 2015, 2016, 2021; Waldner 2015). We stress the importance of putting interview protocols (when interviewing is deemed suitable) at the service of unearthing evidence for each component part of the causal process, including how each step is (causally) connected to the next. Having said this, we acknowledge that in some Bayesian approaches to PT (e.g., Fairfield and Charman 2017, 2022) the researcher doesn't need to systematically trace an entire process to make valid inferences, and evidence doesn't need to be assessed at the level of each component of the causal chain (especially if hypotheses don't make divergent predictions for some components). Finding highly supporting evidence at the level of the general hypothesis or about one component of the chain where the theory makes divergent predictions could increase the posterior probability enough to confirm the advantage of one hypothesis relative to its rivals. While this debate goes beyond the scope of our paper, as noted below, we make it explicit when we think our recommendations suit one view of PT more than the other.

11. Examples include work on case selection (Beach and Pedersen 2018; Saylor 2018), missing data (González-Ocantos and LaPorte 2021), dealing with nonmutually exclusive rival explanations (Rohlfing 2014; Zaks 2017), and combining PT with other methods (Beach 2017).

12. Different classifications have been proposed. For example, Bennett, Fairfield, and Soifer (2019) differentiate between narrative-based PT, Bayesian PT, and mechanistic PT; Runhardt (2021) between "systems" and "interventions" views of PT; and Zaks (2017) between analytic narrative, crisp-set-theoretic, and Bayesian approaches. These classifications do not always map neatly onto each other, and there are divisions within some individual types (e.g., explicit and heuristic Bayesian PT).

13. While PT scholars agree on the importance of causal mechanisms for explanation, there are important disagreements about what causal mechanisms are, how to corroborate them, and their role in inference. See, in particular, a symposium edited by Jacobs (2016) and Mahoney (2001), where he identifies over 20 different definitions of "causal mechanisms." We don't need to take part in this debate to argue that interviews are often a crucial source of data for uncovering evidence about mechanisms. This holds regardless of whether inferences require counterfactuals, productive accounts, or comparing rival hypotheses in a Bayesian framework.

14. This statement might come across as narrow in some respects, as scholars working on mechanisms have also identified exclusively environmental mechanisms (i.e., applied not to actors but their settings) (McAdam, Tarrow, and Tilly 2001; Tilly 2001), or mechanisms that do not involve the actions and interactions of individual or collective actors. While these types of mechanisms can certainly be relevant in bringing about social and political outcomes and, therefore, should be

considered in PT research, interviewing might not be the best technique to uncover evidence for their presence and operation.

15. By being "populated by actors" we don't mean that actors need to be the unit of analysis, just that actors engaging in activities are part of the theorized process. As long as actors' motives for action, or their perceptions and assigned meanings, are important for the process being traced, talking and listening to these actors will be relevant. One can think of this as a scope condition for the recommendations we make here.

16. See, among others, Bernard (2017, Chapter 7); Gerson and Damaske (2020, Chapter 3); Kapiszewski, Maclean, and Read (2015: 212-14); Lynch (2013); Small (2009); Spradley (2016: 25-54); Weiss (1995, Chapter 2).

17. For a discussion of these issues in sociological research, see Small (2009).

18. Small and Calarco (2022: 18-20) stress that "exposure" is what really matters for high-quality qualitative research and highlight that in interview research exposure is not derived from how many people an interviewer talked to, but from the number of hours spent talking to respondents. While spending more time speaking with a relevant subject is likely to yield better interview data, for the specific case of PT we stress relevance before exposure.

19. We embrace the view that interviewing can in principle be used for identifying or uncovering facts about the past. In other words, the data obtained through this method is not exclusively a relational construction between interviewer and interviewee. Data can be *collected* as much as it can be *generated*. For a relational approach to interviewing, see Fujii (2017).

20. Note that a large pool of interviewees does not necessarily say much about representativeness. For example, single-stream snowball sampling can lead us to many people with the same party identification, religious beliefs, level of education, or/and ethnicity, etc. To be sure, we might develop "fingerprints" for which representativeness is important, such as the public opinion data Bateson (2021) uses to cast doubt on alternative explanations in her PT study of Rios Montt's electoral success in Guatemala. However, Bateson does not rely on interview data for this part of her argument. Remember that the point of departure in this discussion is that the components of the theoretical model, and assumptions about their fingerprints, provide guidance on when interview data is an appropriate source.

21. This take on sampling is largely consistent with the positional and reputational criteria for purposive sampling advocated by Tansey (2007: 770-71). However, we stress that in PT these criteria are strongly determined by the theoretical model.

22. Temporal distance is also important in terms of dealing with some of the limitations and challenges of interviews mentioned above: recall error or issues related with memory (Small and Cook 2023; Viterna 2006; Wood 2003, Chapter 2).

23. In addition to evaluating the probative value of testimonial observations, we also need to assess the evidentiary value of testimonies. When we find interview evidence of a posited fingerprint, we should ask whether we can trust the source and, especially when we do *not* find evidence, we should ask whether we interviewed all the relevant actors (i.e., whether we had access to the entire empirical record). Beach and Pedersen (2019: 155-58; Chapter 6) refer to this as evaluating certainty and uniqueness at the empirical level.

24. This approach is similar to what Small (2009: 25-28) calls "sequential interviewing," which he identifies as particularly relevant for studies about processes we do not know much about when we start.

    This recommendation should not be interpreted too rigidly. Researchers could break down the causal chain into different mini-sequences and interview sequentially within each mini-sequence but not overall. They may also choose to start by documenting the outcome and then proceed backward to retrace the steps leading to it—following a reverse sequence. Importantly, sequential interviewing is not always practically possible. For example, we often don't get to choose when actors agree to talk to us and many times we need to take on the opportunity when it appears.

25. Yet, when following this procedure, researchers should be aware of the risk of forming unconscious biases (sequentially adding sets of blinders) that distort data collection as they move on to documenting subsequent component parts of the causal chain.

26. In some variants of explicit Bayesian PT, these evidence tests are deemed unnecessary because inference is dictated by Bayes' rule and probative value is strictly a matter of degree. Yet, this framework also stresses that what matters is not how many empirical observations line up with our theory, but the relative likelihood of the evidence under rival theories (Fairfield and Charman 2017).

27. This logic is consistent with the analytic significance of identifying "costly" or "risky" to adjudicate among competing explanations. See Ingram (2015).

28. Judging which interview statements require to be examined further or be verified by other respondents implies assessing the *certainty* (i.e., the extent to which we *have* to find the posited observable in a given interview) and uniqueness (i.e., the extent to which there are alternatives explanations for the presence of the observable in a given interview) of each fingerprint for each respondent. Doing so is not an easy task. However, the researcher could use detailed knowledge of the context, the identity of the interviewee, and her position in the process, to generate informed and context-sensitive expectations about how certain and unique it is for a specific interviewee to reveal x in the context of the particular case. Answers to these questions should give the interviewer a good sense of whether she needs to

examine further or verify key pieces of interview evidence. See Beach and Pedersen (2019, Chapters 5-7) for more on certainty and uniqueness.

29. For an application of explicit Bayesian PT to this case, see Fairfield and Charman (2017, Appendix, 2022, Chapter 4).

30. Note that stressing that one could stop interviewing representatives of the same "class of actors" doesn't mean that we should stop interviewing other relevant classes of actors, let alone consulting other sources.

31. For more on saturation in interview research, see Fusch and Ness (2015); Guest, Bunce, and Johnson (2006); Small and Calarco (2022: 107-9); Small (2009); and Weller et al. (2018).

32. For example, probability sampling, simple random sampling or different forms of stratified and clustered sampling (Bernard 2017, Chapters 5 and 6; Hibberts, Burke Johnson, and Hudson 2012).

33. Tannenwald's (1999) nuclear taboo offers a classic example of the minimalist approach. Winward's (2021) process linking security forces' low intelligence capacity to mass categorical violence offers an excellent recent example of the more disaggregated form. See Beach and Pedersen (2019: 64-76) for a detailed exposition of these two approaches.

34. For other typologies of mechanisms, see Hedström and Swedberg (1998: 22), Stinchcombe (2005: 170-76) and Beach and Pedersen (2019: 48-51).

35. Moreover, as not every relevant respondent will be in the position to provide valuable information about each component of the model, some interviews will place emphasis on some batteries of questions and omit others depending on the interviewee. This implies that one might need to prepare multiple interview protocols. One time-saving strategy is to create questions for all the components for which interviewing is relevant and then develop templates for each class of interviewees.

36. Lareau (2021: 98) notes that the researcher should ideally find that "focus", and understand better the main priorities, by the time one has conducted about one-half to two-thirds of the interviews. While this sounds about right for interviewing in general, when interviewing in the context of theory-testing PT, this should happen at an earlier stage of the research process.

37. This is particularly useful in the not-uncommon situation when an interviewer is unable to get through the whole interview. This theory-informed system of priorities will tell us which questions are a must. In this situation, if possible, it might be a good idea to prioritize entire components of the process (probably captured in one full battery of questions) for which the interviewee was a crucial actor, even if the interviewee might have something relevant to say about other components.

38. Waldner (2014) makes a compelling case for standardizing the use among process tracers of causal graphs (or directed acyclic graphs) composed of nodes and directed edges connecting nodes and event-history maps, comprising concrete events

that instantiate the graph in the contextual specificity of the case(s) under analysis. The practice of graphically representing process theories can help researchers clearly identify the constituent parts of a process and the form each component could take in an empirical case. It can also prove helpful to identify the components that call for interview data and the evidence that should be looked for.

39. Runhardt (2016) uses Bakke's work to show that focusing on CPOs is not the best way forward for establishing productive continuity and proposes the use of counterfactuals. We do not take part in this debate here, we just want to show how the need to corroborate the different parts of the causal process, including continuity between parts, gives us an indication of the components of the process that interviews might need to cover and type of questions we might need to ask.

40. This is particularly relevant in narrative PT approaches, in approaches that combine PT and historical explanations, as well as in PT approaches in which mechanisms are seen as systems (Beach 2017; Beach and Pedersen 2019; Machamer, Darden, and Craver 2000; Mahoney 2015).

41. Similarly, Bennett and Checkel (2014b: 30) define the observable implications of mechanisms as "the facts and sequences within a case that should be true if each of the alternative hypothesized explanations of the case is true." In discussing the links between PT and historical analysis, Mahoney (2015: 205) also stresses that sequential analysis is inherently linked to process tracing.

42. In explicit Bayesian PT, incorporating information in different but logically equivalent ways (i.e., learning the same pieces of information in different orders) must produce identical probabilities (Bennett 2014: 293; Fairfield and Charman 2019: 157). This doesn't mean, however, that sequence is not important. There is an important difference between learning the same information in different orders and learning about the order (or sequence) of events. Temporal arguments and temporal evidence matter for inference in this tradition (Fairfield and Charman 2022, Chapter 1).

43. Grand tour questions in PT will look more like what the ethnographic interviewing literature calls "specific grand tour question." In this literature, specificity aims to make it easier for the respondent to describe freely, as some might find it difficult to ramble on an abstract question. Therefore, instead of asking "Could you walk me through a normal day in the village?," one asks "Could you walk me through your day yesterday?" (Spradley 2016: 92-106). Being concrete should also help respondents in PT interviewing, however, here specificity should be dictated by the process one is tracing.

44. Process tracers sometimes reconstruct processes by moving backward from the outcome. Structuring questions in this order could also be a good strategy to embed "sequence" into the questionnaire design. If proceeding this way, it is important to be aware that starting with a discussion of the backend components

of a process can "contaminate" participants' accounts of earlier parts, as they might try to adjust their answers about earlier parts to better align with what they said about later ones. Sequencing questions from the beginning might help reduce the temptation to "sanitize" testimonies to avoid inconsistencies and discordances, a common problem in interviews (Gerson and Damaske 2020: 82-83, 90).

45. Question 1: "Thinking about the "truth trials" of the 1990s, which were the main obstacles you encountered when trying to deal with these cases?"

46. Question 2: "I've spoken to judges and prosecutors in [this and other countries] who dealt with similar cases, and they tell me that they were never really trained to deal with juridical idiosyncrasies of these human rights cases, especially the problem of amnesties, statutes of limitations, and double jeopardy. I would like to know more about your experience. (A) What was the "doctrinal consensus" in the judiciary at the time? (B) Do you recall an instance in which you thought it would not be possible to make progress? (C) In interviews with other judges, they also explain that they felt like they were experimenting with the law, interpreting the legality principle in a highly unorthodox way, using arguments derived from international law, etc. Do you agree?."

47. Q3: "How did you become persuaded of the validity and viability of these arguments?"

48. Q4: "I've had access to a number of files that indicate that you attended several seminars organized by (NGO name), and that you eventually became an instructor in those seminars. (A) Why did you attend those seminars? (B) What was the goal of those seminars? (C) How important were these efforts for the judicialization process?"

49. The ability to effectively follow up is an indication of high-quality interview research in general (Small and Calarco 2022, Chapter 4) and, as such, should be so for interviewing for PT.

50. Engaging with alternative explanations is a hallmark of PT in almost all the different approaches. One noticeable exception is Beach and Pedersen (2019: 41-44; Chapter 5).

51. Note that this might also have implications for sampling. If interviewees are selected based on a process theory, then the sample would likely exclude relevant interviewees who might offer essential data to support a rival explanation. In other words, not always we can test a core theory and its rivals with the same interview sample. To truly give rival explanations a real chance, on occasions, we might need an auxiliary sample of respondents.

52. For a detailed treatment of different types of relationships between alternative explanations, how to identify them, and their implications for inference, see Zaks (2017). This issue has been debated among PT methodologists, as some

proponents of the Bayesian approach advocate always articulating rival hypotheses as mutually exclusive. See the exchange between Zaks (2017, 2021, 2022) and Bennett, Fairfield, and Charman (Bennett, Charman, and Fairfield 2022; Fairfield and Charman 2017, 2022, Chapter 7). We don't take part in this debate but recognizing that many process tracers do not articulate rival hypothesis as mutually exclusive, we emphasize the implications the role of alternative hypotheses has for questionnaire design on both scenarios.

53. Leading questions can be useful for tackling alternative explanations. If an alternative account fails to pass a test involving a question phrased in a way that "demands" a confirmatory response, we have stronger reasons for discarding it.

54. In the example from González-Ocantos (2016), the initial open question allowed one to inhabit the world of alternative explanations because it didn't direct the interviewee toward a particular kind of response and was not crafted exclusively to inhabit the world of the proposed explanation. In a world in which pressures from pro-human rights politicians rather than NGOs' resocialization strategies induced a change in judicial behavior, the response given by the judge is highly unexpected. By contrast, it fits the world of the author's theory quite well. Had the question been directed towards the role of NGOs in the process, however, the same response would be less informative in adjudicating between the hypothesis and its rival. This approach is close to the "grand tour" questions discussed above. An alternative approach is to mentally inhabit the world of each hypothesis and ask questions that could elicit answers consistent with each world. See footnotes 43–45 for examples.

55. Incidentally, this points to the potential utility of qualitative pre-analysis plans (PAP-Q) for researchers planning to use PT (Jacobs 2020; Haven et al. 2020). Specifying the causal chain in a pre-analysis plan could be useful as the researcher makes plans for data collection. Writing out the hypothesized causal chain and sketching clearly its observable implications in a PAP-Q might facilitate the process of identifying where interviews are most useful (coverage), who are the key actors (who to interview), and the key activities and their sequence (what to ask and how).

56. See, for example, van Meegdenburg (2023).

References

Anria, Santiago. 2018. *When Movements Become Parties: The Bolivian MAS in Comparative Perspective*. Cambridge: Cambridge University Press.

Baert, Patrick. 1998. *Social Theory in the Twentieth Century*. Cambridge: Polity Press.

Bakke, Kristin M. 2013. "Copying and Learning from Outsiders? Assessing Diffusion from Transnational Insurgents in the Chechen Wars." Pp. 31-62 in *Transnational*

Dynamics of Civil War, edited by Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Bateson, Regina. 2021. "Voting for a Killer: Efraín Ríos Montt's Return to Politics in Democratic Guatemala." *Comparative Politics* 54(2):203-228.

Beach, Derek. 2017. "Process Tracing Methods in the Social Sciences." in *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press.

Beach, Derek and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor: University of Michigan Press.

Beach, Derek and Rasmus Brun Pedersen. 2018. "Selecting Appropriate Cases When Tracing Causal Mechanisms." *Sociological Methods & Research* 47(4):837-71.

Beach, Derek and Rasmus Brun Pedersen. 2019. *Process-Tracing Methods: Foundations and Guidelines*. 2nd ed. Ann Arbor: University of Michigan Press.

Bennett, Andrew. 2010. "Process Tracing and Causal Inference." Pp. 207-20 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier. Lanham, MD: Rowman & Littlefield Publishers.

Bennett, Andrew. 2014. "Appendix: Disciplining Our Conjectures: Systematizing Process Tracing with Bayesian Analysis." Pp. 276-98 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Bennett, Andrew. 2016. "Do New Accounts of Causal Mechanisms Offer Practical Advice for Process Tracing?" *Qualitative & Multi-Method Research* 14(1/2): 34-9.

Bennett, Andrew, Andrew E. Charman, and Tasha Fairfield. 2022. "Understanding Bayesianism: fundamentals for Process Tracers." *Political Analysis* 30(2):298-305.

Bennett, Andrew and Jeffrey T. Checkel (eds) 2014a. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Bennett, Andrew and Jeffrey T. Checkel 2014b. "Process Tracing: From Philosophical Roots to Best Practices." Pp. 3-37 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Bennett, Andrew, Tasha Fairfield, and Hillel David Soifer. 2019. *Comparative Methods and Process Tracing*. APSA - Organized Section for Qualitative and Multi-Method Research. Final Reports.

Bernard, H. Russell. 2017. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. 5th ed. Lanham: Rowman & Littlefield Publishers.

Berry, Jeffrey M. 2002. "Validity and Reliability Issues in Elite Interviewing." *Political Science & Politics* 35(04):679-82.

Bhaskar, Roy. 1978. *A Realist Theory of Science*. Brighton: Harvester Press.

Brady, Henry E. 2010. "Data-Set Observations versus Causal-Process Observations: The 2000 U.S. Presidential Election." Pp. 237-42 in *Rethinking Social Inquiry:*

Diverse Tools, Shared Standards, edited by Henry E. Brady and David Collier. Lanham, MD: Rowman & Littlefield Publishers.

Brady, Henry E. and David Collier (eds) 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 2nd ed. Lanham, MD: Rowman & Littlefield Publishers.

Brady, Henry E., David Collier, and Jason Seawright. 2006. "Toward a Pluralistic Vision of Methodology." *Political Analysis* 14(3):353-68.

Bunge, Mario. 2004. "How Does It Work? The Search for Explanatory Mechanisms." *Philosophy of the Social Sciences* 34(2):182-210.

Campbell, Susanna P. 2018. *Global Governance and Local Peace*. Cambridge: Cambridge University Press.

Chan Tack, Anjanette and Mario L. Small. 2017. "Making Friends in Violent Neighborhoods: Strategies among Elementary School Children." *Sociological Science* 4:224-48.

Clarke, Killian. 2018. "When Do the Dispossessed Protest? Informal Leadership and Mobilization in Syrian Refugee Camps." *Perspectives on Politics* 16(3):617-33.

Closa, Carlos and Stefano Palestini. 2018. "Tutelage and Regime Survival in Regional Organizations' Democracy Protection: The Case of MERCOSUR and UNASUR." *World Politics* 70(3):443-76.

Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44(4):823-30.

Cramer, Katherine. 2016. *The Politics of Resentment*. Chicago: Chicago University Press.

Della Porta, Donatella and Sidney G. Tarrow (eds) 2005. *Transnational Protest and Global Activism*. Lanham, MD: Rowman & Littlefield.

DeLuca, Stefanie, Susan Clampet-Lundquist, and Kathryn Edin. 2016. *Coming of Age in the Other America*. 1st ed. New York: Russell Sage Foundation.

Dunning, Thad. 2014. "Improving Process Tracing: The Case of Multi-Method Research." Pp. 211-36 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Fairfield, Tasha. 2013. "Going Where the Money Is: Strategies for Taxing Economic Elites in Unequal Democracies." *World Development* 47:42-57.

Fairfield, Tasha. 2015a. *Private Wealth and Public Revenue in Latin America: Business Power and Tax Politics*. New York, NY: Cambridge University Press.

Fairfield, Tasha. 2015b. "Reflections on Analytic Transparency in Process Tracing Research." *Qualitative & Multi-Method Research* 13(1):47-51.

Fairfield, Tasha and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3):363-80.

Fairfield, Tasha and Andrew Charman. 2019. "A Dialogue With the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17(1):154-67.

Fairfield, Tasha and Andrew E. Charman. 2022. *Social Inquiry and Bayesian Inference: Rethinking Qualitative Research*. New York: Cambridge University Press.

Fairfield, Tasha and Candelaria Garay. 2017. "Redistribution Under the Right in Latin America: Electoral Competition and Organized Actors in Policymaking." *Comparative Political Studies* 50(14):1871-1906.

Foddy, William. 1993. *Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research*. Cambridge: Cambridge University Press.

Fu, Diana. 2017. *Mobilizing Without the Masses: Control and Contention in China*. Cambridge: Cambridge University Press.

Fujii, Lee Ann. 2017. *Interviewing in Social Science Research: A Relational Approach*. New York: Routledge.

Fusch, Patricia and Lawrence Ness. 2015. "Are We There Yet? Data Saturation in Qualitative Research." *The Qualitative Report* 20(9):1408-16.

George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: The MIT Press.

Gerring, John. 2007. *Case Study Research: Principles and Practices*. New York: Cambridge University Press.

Gerson, Kathleen and Sarah Damaske. 2020. *The Science and Art of Interviewing*. New York: Oxford University Press.

Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

González-Ocantos, Ezequiel A. 2016. *Shifting Legal Visions: Judicial Change and Human Rights Trials in Latin America*. New York: Cambridge University Press.

Gonzalez-Ocantos, Ezequiel and Jody LaPorte. 2021. "Process Tracing and the Problem of Missing Data." *Sociological Methods & Research* 50(3):1407-35.

Guest, Greg, Arwen Bunce, and Laura Johnson. 2006. "How Many Interviews Are Enough?: An Experiment With Data Saturation and Variability." *Field Methods* 18(1):59-82.

Hall, Peter A. 2008. "Systematic Process Analysis: When and How to Use It." *European Political Science* 7(3):304-17.

Handlin, Samuel. 2017. *State Crisis in Fragile Democracies: Polarization and Political Regimes in South America*. New York: Cambridge University Press.

Haven, Tamarinde L., Timothy M. Errington, Kristian Skrede Gleditsch, Leonie van Grootel, Alan M. Jacobs, Florian G. Kern, Rafael Piñeiro, Fernando Rosenblatt, and Lidwine B. Mokkink. 2020. "Preregistering Qualitative Research: A Delphi Study." *International Journal of Qualitative Methods* 19:1-13.

Hedström, Peter and Richard Swedberg (eds) 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.

Hibberts, Mary, R. Burke Johnson, and Kenneth Hudson. 2012. "Common Survey Sampling Techniques." Pp. 53-74 in *Handbook of Survey Methodology for the Social Sciences*, edited by Lior Gideon. New York: Springer.

Holland, Alisha C. 2017. *Forbearance as Redistribution: The Politics of Informal Welfare in Latin America*. Cambridge, New York, UK: Cambridge University Press.

Holmes, Marcus. 2018. *Face-to-Face Diplomacy. Social Neuroscience and International Relations*. Cambridge: Cambridge University Press.

Humphreys, Macartan and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109(4):653-73.

Idler, Annette. 2020. "The Logic of Illicit Flows in Armed Conflict: Explaining Variation in Violent Nonstate Group Interactions in Colombia." *World Politics* 72(3):335-76.

Ingram, Matthew C. 2015. *Crafting Courts in New Democracies: The Politics of Subnational Judicial Reform in Brazil and Mexico*. Cambridge: Cambridge University Press.

Jacobs, Alan M. 2011. *Governing for the Long Term: Democracy and the Politics of Investment*. New York: Cambridge University Press.

Jacobs, Alan M. (ed.) 2016. "Symposium: Causal Mechanisms and Process Tracing." *Qualitative & Multi-Method Research* 14(1/2):13-39.

Jacobs, Alan M. 2020. "Pre-Registration and Results-Free Review in Observational and Qualitative Research." Pp. 221-64 in *The Production of Knowledge. Enhancing Progress in Social Science*, edited by Colin Elman, John Gerring, and James Mahoney. New York: Cambridge University Press.

Kapiszewski, Diana, Lauren M. Maclean, and Benjamin L. Read. 2015. *Field Research in Political Science. Practices and Principles*. New York: Cambridge University Press.

Kaplan, Oliver. 2017. *Resisting War. How Communities Protect Themselves*. Cambridge: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Knott, Eleanor, Aliya Hamid Rao, Kate Summers, and Chana Teeger. 2022. "Interviews in the Social Sciences." *Nature Reviews Methods Primers* 2(1): 1-15.

Kramer, Mark, Bruce J. Allyn, James G. Blight, and David A. Welch. 1990. "Remembering the Cuban Missile Crisis: should We Swallow Oral History?" *International Security* 15(1):212-18.

Lake, Milli May. 2018. *Strong NGOs and Weak States: Pursuing Gender Justice in the Democratic Republic of Congo and South Africa*. New York: Cambridge University Press.

Lareau, Annette. 2021. *Listening to People: A Practical Guide to Interviewing, Participant Observation, Data Analysis, and Writing It All Up*. Chicago: University of Chicago Press.

Lareau, Annette, Elliot Weininger, and Amanda Cox. 2018. "Parental Challenges to Organizational Authority in an Elite School District: The Role of Cultural, Social, and Symbolic Capital." *Teachers College Record* 120(1):1-46.

Leech, Beth L. 2002. "Asking Questions: Techniques for Semistructured Interviews." *PS: Political Science & Politics* 35(4):665-68.

Löblová, Olga. 2018. "When Epistemic Communities Fail: Exploring the Mechanism of Policy Influence." *Policy Studies Journal* 46(1):160-89.

Lyall, Jason. 2014. "Process Tracing, Causal Inference, and Civil War." Pp. 186-208 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Lynch, Julia F. 2013. "Aligning Sampling Strategies with Analytic Goals." Pp. 31-44 in *Interview Research in Political Science*, edited by Layna Mosley. New York: Cornell University Press.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67(1):1-25.

Mahoney, James. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16(3):575-93.

Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41(4):570-97.

Mahoney, James. 2015. "Process Tracing and Historical Explanation." *Security Studies* 24(2):200-18.

Mahoney, James and Rodrigo Barrenechea. 2019. "The Logic of Counterfactual Analysis in Case-Study Explanation." *The British Journal of Sociology* 70(1):306-38.

Mahoney, James and Rachel Sweet Vanderpoel. 2015. "Set Diagrams and Qualitative Research." *Comparative Political Studies* 48(1):65-100.

Masullo, Juan. 2017. "A Theory of Civilian Noncooperation with Armed Groups. Civilian Agency and Self-Protection in the Colombian Civil War." Doctoral Dissertation, European University Institute.

Masullo, Juan. 2021. "Civilian Contention in Civil War: How Ideational Factors Shape Community Responses to Armed Groups." *Comparative Political Studies* 54(10):1849-84.

McAdam, Doug, Sidney G. Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.

Mitchell, Christopher R. 2007. "The Theory and Practice of Sanctuary. From Asylia to Local Zones of Peace." Pp. 1-18 in *Zones of Peace*, edited by Landon E. Hancock and Christopher R. Mitchell. Bloomfield, CT: Kumarian Press.

Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science & Politics* 43(01):29-35.

Mosley, Layna (eds) 2013. *Interview Research in Political Science*. New York: Cornell University Press.

Odell, John. 2006. "A Major Milestone With One Major Limitation." *Qualitative Methods: Newsletter of the APSA Qualitative Methods Section* 4(1):37-40.

O'Donnell, Guillermo, Philippe C. Schmitter, and Laurence Whitehead. 1986. *Transitions from Authoritarian Rule: Comparative Perspectives*. Baltimore: John Hopkins University Press.

Pouliot, Vincent. 2014. "Practice Tracing." Pp. 237-58 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Ricks, Jacob I. and Amy H. Liu. 2018. "Process-Tracing Research Designs: A Practical Guide." *PS: Political Science & Politics* 51(4):842-46.

Rohlfing, Ingo. 2014. "Comparative Hypothesis Testing Via Process Tracing." *Sociological Methods & Research* 43(4):606-42.

Runhardt, Rosa W. 2015. "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation." *Philosophy of Science* 82(5):1296-307.

Runhardt, Rosa W. 2016. "Tracing the Productive Continuity of Social Mechanisms." *Qualitative & Multi-Method Research* 14(1/2):22-8.

Runhardt, Rosa W. 2021. "Evidential Pluralism and Epistemic Reliability in Political Science: Deciphering Contradictions Between Process Tracing Methodologies." *Philosophy of the Social Sciences* 51(4):425-42.

Saylor, Ryan. 2018. "Why Causal Mechanisms and Process Tracing Should Alter Case Selection Guidance." *Sociological Methods & Research* 49(4): 982-1017.

Schimmelfennig, Frank. 2003. *The EU, NATO and the Integration of Europe: Rules and Rhetoric*. Cambridge: Cambridge University Press.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.

Small, Mario L. 2009. "'How Many Cases Do I Need?': On Science and the Logic of Case Selection in Field-Based Research." *Ethnography* 10(1):5-38.

Small, Mario L. and Jessica Mccrory Calarco. 2022. *Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research*. Oakland: University of California Press.

Small, Mario L. and Jenna M. Cook. 2023. "Using Interviews to Understand Why: Challenges and Strategies in the Study of Motivated Action." *Sociological Methods & Research* 52(4):1591-631.

Sober, Elliott. 2009. "Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-Tuning, and Firing Squads." *Philosophical Studies* 143(1):63-90.

Spradley, James P. 2016. *The Ethnographic Interview*. Long Grove: Waveland Press.

Stinchcombe, Arthur L. 2005. *The Logic of Social Research*. Chicago: University of Chicago Press.

Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3):433-68.

Tannenwald, Nina. 2007. *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945*. Cambridge: Cambridge University Press.

Tansey, Oisín. 2007. "Process Tracing and Elite Interviewing: A Case for Non-Probability Sampling." *PS: Political Science & Politics* 40(04):765-72.

Tavory, Iddo. 2020. "Interviews and Inference: Making Sense of Interview Data in Qualitative Research." *Qualitative Sociology* 43(4):449-65.

Tilly, Charles. 2001. "Mechanisms in Political Processes." *Annual Review of Political Science* 4(1):21-41.

Van Meegdenburg, Hilde. 2023. "Process Tracing: An Analyticist Approach." Pp. 405-20 in *Routledge Handbook of Foreign Policy Analysis Methods*. London: Routledge.

Viterna, Jocelyn. 2006. "Pulled, Pushed, and Persuaded: Explaining Women's Mobilization into the Salvadoran Guerrilla Army." *American Journal of Sociology* 112(1):1-45.

Waldner, David. 2014. "What Makes Process Tracing Good? Causal Mechanisms, Causal Inference, and the Completeness Standard in Comparative Politics." Pp. 126-52 in *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge: Cambridge University Press.

Waldner, David. 2015. "Process Tracing and Qualitative Causal Inference." *Security Studies* 24(2):239-50.

Weiss, Robert S. 1995. *Learning From Strangers: The Art and Method of Qualitative Interview Studies*. New York: Free Press.

Weller, Susan C., et al. 2018. "Open-Ended Interview Questions and Saturation." *PLoS ONE* 13(6):1-18.

Winward, Mark. 2021. "Intelligence Capacity and Mass Violence: Evidence From Indonesia." *Comparative Political Studies* 54(3–4):553-84.

Wood, Elisabeth J. 2000. *Forging Democracy From Below: Insurgent Transitions in South Africa and El Salvador*. New York: Cambridge University Press.

Wood, Elisabeth J. 2001. "An Insurgent Path to Democracy: Popular Mobilization, Economic Interests, and Regime Transition in South Africa and El Salvador." *Comparative Political Studies* 34(8):862-88.

Wood, Elisabeth J. 2003. *Insurgent Collective Action and Civil War in El Salvador.* Cambridge: Cambridge University Press.

Zaks, Sherry. 2017. "Relationships Among Rivals (RAR): A Framework for Analyzing Contending Hypotheses in Process Tracing." *Political Analysis* 25(3):344-62.

Zaks, Sherry. 2021. "Updating Bayesian(s): A Critical Evaluation of Bayesian Process Tracing." *Political Analysis* 29(1):58-74.

Zaks, Sherry. 2022. "Return to the Scene of the Crime: Revisiting Process Tracing, Bayesianism, and Murder." *Political Analysis* 30(2):306-10.

Author Biographies

Ezequiel Gonzalez-Ocantos is a Professor of Comparative & Judicial Politics in the Department of Politics and International Relations at the University of Oxford, and Professorial Fellow of Nuffield College. He is the author of *Shifting Legal Visions: Judicial Change and Human Rights Trials in Latin America* (CUP 2016), *The Politics of Transitional Justice in Latin America* (CUP 2020), and *Prosecutors, Voters, and the Criminalisation of Corruption in Latin America* (CUP 2023), and also co-editor of *The Limits of Judicialization: From Progress to Backlash in Latin America* (CUP 2023). Currently, he co-edits Qualitative and Mixed-Methods Research (QMMR), a biannual publication of the American Political Science Association.

Juan Masullo is an Assistant Professor (with tenure) of Political Science at the Institute of Political Science, Leiden University, and a member of the Amsterdam Conflict Research Network. His research employs a variety of methods to study civil wars and organized crime, especially in Latin America and Italy. He has published on these topics in several recognized journals and recentlly co-edited the volume *Civilian Protective Agency in Violent Settings* (OUP 2023). He is co-editor of Qualitative and Mixed-Methods Research (QMMR), a biannual publication of the American Political Science Association.