Tasha Fairfield and Andrew Charman

Explicit Bayesian analysis for process tracing: guidelines, opportunities, and caveats

Article (Accepted version)
(Refereed)

Explicit Bayesian Analysis for Process Tracing:
Guidelines, Opportunities, and Caveats

V3.5: January 2017

Tasha Fairfield
Dept. of International Development
London School of Economics
T.A.Fairfield@lse.ac.uk

Andrew Charman
Department of Physics
University of California, Berkeley

Abstract

Bayesian probability holds the potential to serve as an important bridge between qualitative and quantitative methodology. Yet whereas Bayesian statistical techniques have been successfully elaborated for quantitative research, applying Bayesian probability to qualitative research remains an open frontier. This paper advances the burgeoning literature on Bayesian process tracing by drawing on expositions of Bayesian "probability as extended logic" from the physical sciences, where probabilities represent rational degrees of belief in propositions given the inevitably limited information we possess. We provide step-by-step guidelines for explicit Bayesian process tracing, calling attention to technical points that have been overlooked or inadequately addressed, and we illustrate how to apply this approach with the first systematic application to a case study that draws on multiple pieces of detailed evidence. While we caution that efforts to explicitly apply Bayesian learning in qualitative social science will inevitably run up against the difficulty that probabilities cannot be unambiguously specified, we nevertheless envision important roles for explicit Bayesian analysis in pinpointing the locus of contention when scholars disagree on inferences, and in training intuition to follow Bayesian probability more systematically.

1. Introduction

A growing movement within political science has identified Bayesianism as the methodological foundation of process tracing, which entails making causal inferences about a single case by assessing alternative explanations in light of evidence uncovered.[1] Compared to frequentism, Bayesianism offers several advantages that are especially important for qualitative case research. Bayesian probability can handle data that are not generated from a stochastic process (e.g., information from expert informants and archival sources), it can be applied to explain unique events without reference to a population, and it mandates an iterative "dialogue with the data,"[2] which mirrors how process tracing is usually conducted.

As part of an initiative to improve analytical transparency and establish process tracing as a rigorous method, the literature has moved from informal analogies to Bayesianism (McKeown 1999, Bennett 2008, Beach and Pedersen 2013) toward efforts to explicitly apply Bayesian analysis in qualitative research (Rohlfing 2013, Bennett 2015, Humphreys and Jacobs 2015). We view this turn to Bayesianism as a watershed that provides solid grounding for in-depth, small-N case research. However, whereas Bayesian statistical techniques have been successfully elaborated for large-N quantitative research (eg. Gill 2008, Jackman 2009, Gelman et al. 2013), applying Bayesian probability in qualitative case research remains a frontier that has not been definitively addressed. To date, scholars have examined only a few illustrative pieces of evidence (Rohlfing 2013, Bennett 2015) and/or include only highly simplified process-tracing clues (Humphreys and Jacobs 2015). Moreover, a number of technical points have been overlooked or inadequately addressed within the literature that is innovating in this terrain.

We aim to advance efforts to apply Bayesian reasoning in process tracing by drawing on expositions of Bayesian "probability as extended logic" that originated in the natural sciences, dating back to Bernouille (1655-1705) and Laplace (1749-1827). Modern physicists (Cox 1961, Jaynes 2003) have demonstrated mathematically that Bayesian probability theory provides the uniquely consistent extension of deductive logic, where we know whether any given proposition is true or false, to situations where information is incomplete, uncertainty reigns, and hypotheses can rarely be definitively proven or disproven. This "logical" Bayesian notion of probability—the rational degree of belief that we should hold in a hypothesis or some other proposition in light of the information we possess—in principle provides a unified framework for inference.

We being by situating this paper's contributions within the burgeoning literature on Bayesian process tracing (Section 2). Section 3 provides a step-by-step exposition of how to explicitly apply Bayes' rule to draw inferences in process tracing research. We emphasize several conceptual and technical points that diverge from previous treatments, including comparing a hypothesis *H* against clearly delineated rivals, rather than its unspecified logical negation *~H,* and using a logarithmic scale instead of a linear scale to assign probabilities, with an analogy to sound, which minimizes arbitrariness and facilitates meaningful assessments of uncertainty. Section 4 provides illustrations from our systematic application of explicit Bayesian analysis to a published case study that draws on multiple pieces of evidence; Appendix A elaborates this case application in detail.

Section 5 concludes by assessing the potential for explicit Bayesian analysis to improve causal inference and analytic transparency in qualitative research, an important and timely issue given ongoing debate within political science on how best to promote research transparency

[1] Bayesianism also underpins casual analysis in qualitative research much more broadly, including assessing higher-level theories in light of multiple cases.

[2] Astrophysicist Stephen Gull, quoted in Sivia (2006).

(www.dartstatement.org, Lupia and Elman 2014, QMMR 2015, www.qualtd.net). We envision important roles for explicit Bayesian analysis in communicating judgments and pinpointing the locus of contention when scholars disagree on inferences, as well as training intuition to make more systematic and logically rigorous conclusions. However, we caution that assigning numerical values to probabilities in qualitative research involves a substantial dose of arbitrariness that can limit the utility of explicitly employing Bayesian analysis when assessing complex evidence and nuanced causal models. These caveats do not undermine the importance of Bayesian probability as the aspirational ideal of scientific inference and the methodological foundation for process tracing. Understanding the technical details of Bayesian probability that we elaborate can help discipline our reasoning and elucidate best practices for process tracing, whether formal or narrative-based.

Beyond providing rigorous analytical tools and insights for process-tracing practitioners, this paper aims to foster greater understanding of the inferential logic that underlies qualitative case research among a broader political science audience. All research, from large-N econometrics to historical analysis, draws on insights from qualitative information, and we believe that Bayesian probability can serve as an important bridge between qualitative and quantitative methodology.

2. Mapping the Intellectual Terrain

Treatments of Bayesianism in political science draw their foundations on what we call the "psychological" school, which underlies most Bayesian epistemology in philosophy of science and Bayesian statistics textbooks (e.g., Savage 1954, Howson and Urbach 2006). In contrast, we advocate "logical" Bayesianism, as articulated in the physical sciences, as the methodological foundation for process tracing and scientific inference more broadly. Whereas psychological Bayesianism treats probabilities as a matter of informed opinion, logical Bayesianism seeks to represent the rational degree of belief that we should hold in propositions given the *information* we possess, independently of whims, hopes, or personal predilections. A probability $P(A|B)$ represents the degree of belief that is rational to hold in proposition A if nothing but B is presumed known.

A central tenet of logical Bayesianism is that probabilities should encode knowledge in a unique, "consistent" manner. Incorporating information in different but logically equivalent ways (e.g. learning the same pieces of information in a different order) must produce identical probabilities, and rational individuals who possess the same information must assign the same probabilities. Cox (1961) and Jaynes (2003) demonstrated that if we demand consistency and measure degrees of belief with real numbers between zero and one, then we must adopt the standard sum and product rules for manipulating and updating probabilities as the unique mode of reasoning in the face of uncertainty. Bayes' theorem and all other mathematical operations in Bayesian analysis follow directly from these basic rules.

Logical Bayesianism is prescriptive, in that it aspires to elaborate procedures that a rational individual *should* follow when reasoning with limited information. In practice, simplifications and approximations are usually required. Even in the natural sciences, the full calculations may be too difficult to complete or even set up, and prior knowledge cannot always be translated unambiguously into prior probabilities. In social science and especially in qualitative research, we must accept some degree of subjectivity as inevitable. But we can nevertheless train our thinking to better approximate logical Bayesianism; this paper aims to further that goal.

Moving from methodological foundations to practice, pioneering research on process tracing has not yet resolved the question of how to operationalize Bayesian reasoning for in-depth case analysis that evaluates complex hypotheses and large amounts of nuanced qualitative information. We build on Bennett (2015) by providing a technical exposition grounded in logical Bayesianism. We expound key principles including conditioning likelihoods on previously-incorporated evidence that are noted but not fully explicated or applied in previous treatments, and we stress the necessity of comparing well-specified rival hypotheses. Abell (2009) shares some similarities with our approach but focuses on inferring the existence of single casual "links" in a narrative and does not recognize the importance of assessing well-posed alternative explanations rather than asking whether a given causal mechanism is present or absent.

Turning to the multi-methods literature, Humphreys and Jacobs (2015) break new ground with their Bayesian model (BIQQ) for combining correlational data with simple process-tracing clues. Their work points to the importance of Bayesian probability as a unified framework of inference. However, if our goal is applying Bayesian analysis to evidence-intensive process tracing, the BIQQ model is more complicated than needed. Drawing on medical testing analogies, Humphreys and Jacobs classify cases into types (adverse, beneficial, chronic, destined) according to the potential outcome a treatment would elicit. Because these types are unobservable and carry no information about causal mechanisms, we regard them as nuisance parameters. This setup becomes cumbersome if we are searching for the best explanation for a given case, rather than assessing average causal effects and other population-level parameters from a sample. Instead of conditioning on the case's hidden type, we advocate directly evaluating the likelihood of the evidence conditional on each hypothesis we wish to compare.

A common feature of both the Bayesian process-tracing literature and the Bayesian multi-methods literature is the central role they retain for Van Evera's (1997) process-tracing tests (Bennett 2008 & 2015, Collier 2011, Mahoney 2012, Humphreys and Jacobs 2015). We view classification of tests as unnecessary within a Bayesian framework, since evidentiary confirmation is always a matter of degree, not type, and inference is always governed by the logic of Bayes' rule. In addition, although much of Van Evera's reasoning is intuitively Bayesian, the notion of process tracing tests remains close in spirit to frequentist statistical theory, with its tradition of subjecting hypotheses to a sequence of discrete, named tests that can be either passed or failed. Instead, we again advocate directly evaluating likelihoods under alternative hypotheses, as explained in the following section.

3. Operationalizing Bayesian Analysis

Explicit Bayesian process tracing involves three key steps: (§3.1) specifying hypotheses H_i and assigning their prior probabilities, $P(H_i|I)$, given relevant background information I, (§3.2) identifying the evidence, E, and (§3.3) assessing likelihoods, $P(E|H_i I)$, and/or likelihood ratios, $P(E|H_i I)/P(E|H_j I)$. We can then obtain the posterior odds on H_i vs. H_j in light of the evidence by applying the relative odds-ratio form of Bayes' rule:

$$\frac{P(H_i|E\, I)}{P(H_j|E\, I)} = \frac{P(H_i|\, I)}{P(H_j|\, I)} \times \frac{P(E|H_i\, I)}{P(E|H_j\, I)}\ , \tag{1}$$

In §3.4, we explicate the rationale for using a logarithmic scale when assigning numerical values to probabilities; §3.5 illustrates how to derive an aggregate inference from multiple pieces of evidence. Looking forward, we stress that evaluating what Bayesian analysis can do for process tracing as well as the challenges and limitations requires an understanding of the technical aspects discussed below.

3.1 Hypotheses and Priors

The first step in explicit Bayesian process tracing entails specifying the set of hypotheses we wish to consider, $\{H_k\}$. Whereas most treatments compare a single working hypothesis H against it logical negation, $\sim H$, we advocate identifying one or more concrete rival hypotheses. This approach is critical in social science, because $\sim H$ generally will not be a well-defined proposition—H could fail to hold in an essentially infinite number of ways. The more specific the hypothesis is, the more possibilities are embodied in $\sim H$. Directly assessing likelihoods of the form $P(E|\sim H\, I)$ will be practically impossible if we have not first contemplated what concrete possibilities $\sim H$ might actually entail. Suppose H_A = *Suspect A killed the victim*, and we have a clue E = *A's glove was found at the crime scene*. If the only plausible alternative hypothesis is $H_B = \sim H_A$ = *Suspect B, A's bitter ex-spouse, killed the victim and framed A*, then $P(E|\sim H_A\, I)$ could be fairly high, assuming the ex-spouse still had some access to *A*'s personal effects. However, if the only plausible alternative hypothesis is $H_C = \sim H_A$ = *Suspect C, a total stranger to A, killed the victim,* then $P(E|\sim H_A\, I)$ might be quite low. If both H_B and H_C are considered plausible alternatives, the only sensible way to proceed is to evaluate the likelihood conditional on each hypothesis separately, $P(E|H_i\, I)$. We can then calculate posterior odds ratios for H_A vs. H_B and H_A vs. H_C, or we can use Bayes' rule to obtain the posterior probability on each of the three hypotheses and then calculate $P(\sim H_A|E\, I) = 1-P(H_A|E\, I)$, keeping in mind that our background information includes the assumption that $\sim H_A = H_B + H_C$.

Whenever possible, mutually exclusive hypotheses are preferable. In principal, we can always take a set of non-rival hypotheses and construct a set of mutually exclusive rivals. For example, consider two factors that could motivate presidents to violate protectionist policy mandates: a desire to represent voters' best interests, and a desire to seek rents associated with neoliberalism (Stokes 2001). If we think both factors may contribute, we could delineate five rivals: H_1 = predominantly representation, H_2 = both but mostly representation, H_3 = both in relatively equal measure, H_4 = both but mostly rent-seeking, H_5 = predominantly rent-seeking. Strictly speaking, ensuring that these possibilities are mutually exclusive requires greater precision—what exactly do we mean by "predominantly" vs. "mostly" vs. "relatively equal"? This specification issue can pose challenges in explicit Bayesian process tracing.[3] In practice, it is important to specify hypotheses as carefully as possible and to explicitly include the assumption that they are mutually exclusive and exhaustive as part of the background information. If new evidence suggests that a more complex hypothesis would provide the best explanation, we should incorporate it in $\{H_k\}$ and redo the analysis.

Once we have specified rival hypotheses, we must assign prior probabilities. In principle, we advocate starting from an initial state of maximal ignorance, I_0, and placing equal prior probabilities on each hypothesis,[4] in accord with the "principle of indifference" or Laplace's principle of "insufficient reason," (Gregory 2005:37-38; Jaynes 2003:40-41). We then build up via Bayes' rule from I_0 to the actual prior state of knowledge I. In practice, especially for social science, systematically incorporating all of our background information is infeasible. There is

[3] Additional complications arise if we wish to model how the two factors contribute in $H_{2\text{–}4}$. They could act independently, or they could interact—representation might serve as a means to the end of long-term rent-seeking through continuity in office, if neoliberalism affords sustained opportunities for corruption.

[4] Assuming a discrete set of mutually-exclusive hypotheses.

simply too much such information to exhaustively list, setting aside the challenges of formally applying Bayes' rule to each such piece of information. Given this reality, there are several reasonable options for proceeding. First, we can specify priors that aim to reflect our background information as best as possible, recognizing that the logical Bayesian approach is aspirational and some subjectivity will inevitably enter in practice. Second, we can simply use equal priors, which avoids biasing the initial assessment in favor of any of the hypotheses. Third, we can report likelihood ratios instead of posterior probabilities and allow readers to supply their own priors. If the evidence is strong, scholars may converge on a single hypothesis even if they start from different priors. If the evidence does not strongly favor a single hypothesis, scholars may at least be able to agree on the direction in which their beliefs should be shifted. Another option entails carrying out the analysis starting from several different prior probability distributions, which allows us to assess how sensitive our conclusions are to these initial choices.

Leading literature on Bayesian process tracing does not always follow these guidelines. Consider Bennett's (2015) discussion of Tannenwald's (2007) research on the non-use of nuclear weapons in the postwar period. He focuses on Tannenwald's three principle alternative hypotheses, which we denote H_D = deterrence, H_M = lack of military utility, and H_T = norms, in the form of a "nuclear taboo." Bennett (2015:277) observes that these hypotheses "at first glance seem equally plausible." In accord with this assessment, which corresponds to the indifference principle, we should use equal prior probabilities of 1/3 for each hypothesis. However, Bennett (2015:278) instead chooses a prior of 40% for H_T and 60% for $\sim H_T$ ($=H_D+H_M$). Rohlfling (2013:13-16) in contrast constructs priors through a process that entails identifying one working hypothesis, assigning a preliminary prior probability of 50%, discovering a rival hypotheses by exploring the literature, and then reducing the prior probability on the working hypothesis by what appears to be an arbitrary amount. After two iterations corresponding to the discovery of two alternative hypotheses, he ultimately gives the working hypothesis a prior of 30%. Instead, best practice mandates that we should state each hypothesis from the outset, before assigning priors. When comparing three mutually exclusive hypotheses, we could then assign equal prior probabilities of 1/3, or we could choose values that aim to reflect our background information, with an explanation of why we favor some hypotheses over others.

Reiterating the critical points, before assigning priors, we must elaborate a clearly articulated set of hypotheses that we assume to be mutually exclusive and exhaustive. We should then assign a probability to each hypothesis, rather than considering only the working hypothesis and its logical negation, which implicitly contains all of the rivals. If we discover or devise a new hypothesis later on, we must start the problem over and reassign priors.

3.2 Evidence

We take a broad, common-sense view of what constitutes evidence in process tracing—any relevant observation or information (beyond our background knowledge) that bears on the truth of our hypotheses. Evidence often contains information about timing and sequencing, actors' goals and intentions, and other aspects of causal mechanisms, as obtained from a wide range of sources including interviews, archives, media records, and secondary literature. Bayesianism does not face any of the restrictions that limit frequentist statistics to analyzing only replicable observations or data generated from some stochastic process, such as random sampling. The type of evidence—regardless of how distinctions are delimited—does not matter for the fundamental logic of Bayesian analysis, because we are allowed to evaluate the probability of any proposition.[5] For example, the evidence we evaluate in our empirical example includes not only "within-case" observations about the causal process driving Chile's 2005 tax reform, but also "cross-case" evidence from previous tax reform episodes that bears on the hypotheses for explaining the 2005 reform (see E_1, Section 4). We view classification of evidence as superfluous unless it helps us evaluate likelihoods, because the evidence enters Bayesian calculations *only* through likelihoods, regardless of its origin or form.

In a Bayesian framework, attempting to provide a definition of what constitutes "good" evidence is largely beside the point.[6] Evidence can be more or less informative, as determined by the likelihood ratio, but probative value—or "weight of evidence"[7]—falls on a continuum. Moreover, what matters is how strongly the total body of evidence discriminates between rival hypotheses. For example, multiple pieces of information that each yield a very small weight of evidence together with one piece of information that produces a large weight of evidence may give an overall inference of essentially the same strength as a single highly probative piece of information, or two pieces of information with intermediate weights of evidence. While we would ideally like every piece of evidence we gather to be highly probative, in social science we must often make inferences based on the accumulated weight of evidence from many clues, none of which is strongly decisive. On the other hand, we sometimes do encounter a highly decisive piece of evidence, in which case we need not spend time systematically assessing other marginally probative pieces of evidence, because the latter will not affect the overall inference.

Likewise, there is no general prescription for designating what constitutes a distinct "piece" of evidence. We are free to disaggregate the overall body of evidence E into components E_1–E_N as finely or as coarsely as we see fit, with the goal of parsing information at a level that facilitates reasoning about likelihoods (Section 3.3). As a broad rule of thumb, observations that intuitively appear to favor different hypotheses, or information derived from distinct types of sources (e.g. a right-wing politician vs. a labor-union leader) are usually best treated as separate pieces of evidence, whereas similar information arising from similar sources (e.g. two government informants tell a similar story about a policy process) might usefully be treated as a single piece of evidence. A second rule of thumb would be to avoid disaggregating the evidence too finely; if we seek to make too many analytical steps explicit, we risk becoming lost in minutia. On the other hand, if we aggregate too much evidence, it becomes difficult to assess the overall likelihood of what is actually a conjunction of many distinct propositions.

3.3 Likelihoods and Likelihood Ratios

Assessing likelihoods and likelihood ratios is the key inferential step in Bayesian process tracing. The likelihood of the evidence tells us how we should update our prior degree of belief in a given hypothesis, and, more importantly, likelihood ratios allow us to adjudicate between rival hypotheses and thereby identify the best explanation in light of the evidence. This section explains how best to interpret and assess likelihoods, with guidelines for handling information from sources that may not be reliable and accounting for logical dependence among multiple pieces of evidence.

[5] The difficulty of assigning probabilities may nevertheless vary.

[6] Aside from issues of data validation (e.g. correctly recording sources, avoiding mistranslation).

[7] §3.4 gives a precise mathematical definition for the weight of evidence.

3.3.1 Inhabit the World of Each Hypothesis

The likelihood $P(E|H_i\, I)$ is our degree of belief in the truth of E—a proposition which states some specific empirical evidence—conditional on a given hypothesis and our background information. Colloquially, we sometimes speak of the probability of 'finding or observing the evidence,' since we are typically interested in updating beliefs in light of evidence that we have actually obtained. When using this language, we implicitly assume that proposition E includes the fact that the researcher learned or reliably observed the relevant information.[8]

The key to assessing $P(E|H_i\, I)$ is to remember that we are assuming H_i is correct. Recall that in the standard notation of conditional probability, everything to the right of the vertical bar is "conditioning information" that is either reliably known or assumed as a conjecture when reasoning about the probability of the proposition to the left of the bar. While we do not actually know whether H_i is true, we must nevertheless consider the implications of H_i being true. In Hunter's (1984) illuminating terminology, we must "mentally inhabit the world of the hypothesis" and ask how surprising (low probability) or expected (high probability) the evidence would be in that world, or in other words, how likely would it be for E to hold true in that world. If the evidence is more probable in the 'H_1 world' relative to the 'H_2 world,' then that evidence will increase the odds we place on H_1 vs. H_2.[9] That is, we gain confidence in one hypothesis to the extent that it makes the evidence we see more plausible in comparison to the rival hypothesis.

Assessing likelihoods entails thinking about how consistent the evidence is with the world of the hypothesis in question and imagining what other kinds of observations or scenarios we might expect in that world. However, we cannot hope to enumerate a full list of possible clues that could have occurred; as Leibniz remarked, "evidence is not be counted but weighed." Section 4 provides examples that illustrate the reasoning process behind assessing likelihoods.

3.3.2 Testimonial Evidence: Assess the likelihood that "source S stated X"

In the natural sciences, a measurement apparatus usually needs to be calibrated before gathering and/or analyzing the data. In the social sciences, we often use "testimonial evidence" (La Place) provided by people (e.g. politicians, journalists, historians) who may misremember, dissemble, or skew facts to suit their interests. The closet analog of calibration might entail assessing the reliability of the source; yet in general we cannot do so in any absolute sense, independently of what exactly was said and which hypothesis holds among the set of alternatives under consideration. Suppose we are studying a pension reform legislated in a developing country, and a close advisor to the president tells us: "Congress was willing to sit down and negotiate after the president explained to the public that 'there will be no money for people to retire in five years if we don't do the reform now.'" We might consider two rival hypotheses: H_1= *the legislation passed because the president made compelling public appeals that put pressure on congress to approve the reform*, and H_2 = *the reform passed because the president's chief of staff bribed members of congress*. In the world of H_1, what the informant has said is true and accurate, whereas in the world of H_2, the informant has provided incorrect information.

[8] In other words, if E = 'The scholar observed E_0' then asking 'what is the probability that E is true given H' is equivalent to asking 'what is the probability of observing E_0 if H is true.' We sometimes elide the extra verbiage in E and use E_0 as a shortcut.

[9] In our usage, 'the world of H' defines a family of possibilities where H holds, but various contingencies consistent with H (which might constitute evidence or clues) may or may not occur. Others might describe these as multiple possible worlds (Barrenechea & Mahoney 2016).

Testimonial evidence in process tracing is best analyzed by including the source in the definition of the evidence. In other words, evidence E should typically take the form "source S stated X." For example, we might have E =*An article in a left-leaning newspaper reported X.* We can then directly evaluate the likelihood of this evidence given a specific hypothesis. Our background information will include assessments of the source's knowledgeability and incentives to reveal or distort the truth under a particular hypothesis, as well as a range of salient contextual clues (e.g. body language and intonation for interviews). Revisiting the pension reform example above and taking E = *A close presidential advisor told us in an interview*: *"Congress was willing to sit down and negotiate after the president explained...,"* we might judge $P(E|H_1\, I\,)$ to be high, because the informant has few incentives to withhold or distort the truth in this world. Under the corruption hypothesis, instrumental incentives could easily compel the informant to tell the (incorrect) story about the president's public appeal, so E is not unexpected, but the value we assign to $P(E|H_2\, I)$ will depend on our rapport and trust in the informant, as informed by our background information.

These guidelines for assessing testimonial evidence follow Jaynes (2003:128), who emphasizes that the new information we learn in both science and politics is that a source has *claimed* some facts, not the purported facts themselves. Contrary to this best practice, Beach and Pedersen (2013:126-29) regard the statement X as the evidence and attempt to assess the probability that X is accurate as an additional, distinct component of the analysis in a way that is hypothesis independent. This approach cannot work, because in general, the accuracy of information X *must* depend on the hypothesis under consideration (see Appendix B for a mathematical elaboration).

3.3.3. Condition on Previously-Incorporated Evidence

When the total body of evidence E consists of multiple observations E_1–E_N, we can decompose the likelihood into a product as follows:

$$\begin{aligned} P(E|H_k\, I) &= P(E_1 E_2 \ldots E_N|H_k\, I) = P\big(E_N\big|E_1 E_2 \ldots E_{(N\text{-}1)}\, H_k\, I\big)\, P\big(E_1 E_2 \ldots E_{(N\text{-}1)}|H_k\, I\big) = \cdots \\ &= P\big(E_N\big|E_1 \ldots E_{(N\text{-}1)}\, H_k\, I\big) \ldots P(E_2\,|E_1\, H_k\, I)\, P(E_1|\, H_k\, I)\ , \end{aligned} \tag{2}$$

because the product rule allows us to write the joint probability of A and B as $P(AB) = P(A|B)*P(B)$. The right-hand side of the equation (2) is essentially a product of likelihoods for each observation E_x, with the nuance that previously-analyzed evidence must be incorporated along with I as conditioning information. In other words, the likelihood for E_x must be assessed conditional not only on a hypothesis and the background information, but also on all evidence from the current problem that we have previously incorporated, E_{prev}. We must therefore ask if E_x is any more or less likely given that we already know E_{prev}, beyond whatever H_k and I imply.

Conditioning on previously-incorporated evidence requires careful thought about logical dependence between E_x and E_{prev} under each hypothesis. *Logical* dependence is distinct from *causal* (or physical) dependence, in that E_{prev} need not exert any causal influence on E_x for it to affect our degree of belief in E_x. Consider a textbook example of drawing without replacement from an urn containing two red marbles and one black marble. Suppose we close our eyes, draw one marble from the urn, and set it aside without looking. We then repeat the process and draw a second marble from the urn. We are interested in the probability that the first draw produced a red marble, $P(E_{1R}\mid I\,)$. Before looking at either marble taken from the urn, $P(E_{1R}\mid I\,) = 2/3$. If we observe the *second* marble drawn from the urn and discover that it is black, we gain information that is highly relevant for inferences about the color of the first marble—in light of this information, we become certain that the first marble is red: $P(E_{1R} \mid E_{2B}\ \ I) = 1$. The outcome of the second draw can in no way exert a causal influence backward in time on the outcome of the first draw; the dependence is entirely logical in nature. Causal interactions between systems or causal influences in their common past can of course lead to logical dependence, but in a Bayesian framework, it is essential to remember that conditional probabilities reflect logical dependencies between propositions, not physical connections.

Returning to social science, suppose an informant interviewed in December 2005 tells a story X, which we will denote as evidence $E_{Inf(X)}$, and a news article from May 2005 reports a similar story, denoted as evidence $E_{News(X)}$. Suppose we conduct and analyze the interview first, and then subsequently discover the news article, such that $E_1 = E_{Inf(X)}$ and $E_2 = E_{News(X)}$. In this case, E_1 cannot have any causal effect on E_2—what the informant has told us cannot possibly influence a news article written seven months earlier. Nevertheless, $P(E_2 \mid E_1\ H\ I)$ will not equal $P(E_2 \mid H\ I)$. Whether $P(E_2 \mid E_1\ H\ I)$ is higher or lower than $P(E_2 \mid H\ I)$ may depend on the hypothesis H. The point is that $E_{News(X)}$ and $E_{Inf(X)}$ are logically dependent given possible causal connections that might have occurred in the past. For instance, our informant may have learned X from reading the news article, so under many hypotheses, we would be less surprised to encounter the article after talking to the informant, in which case $P(E_2 \mid E_1\ H\ I) > P(E_2 \mid H\ I)$.

It is also important to note that dependence (or independence) is not a physical property of our sources—e.g. the newspaper vs. the informant in the example above—notwithstanding what might be suggested by common advice to seek "independent sources of evidence."[10] Instead, it is a logical relationship between pieces of evidence given a specific hypothesis. For any two pieces of evidence—which should include information about the source—it is almost always possible to concoct some hypothesis under which they are dependent. Drawing on different sources and distinct types of information is certainly good practice, but it does not absolve us from thinking carefully about potential logical dependence among the data. The degree to which one source corroborates another depends on the hypothesis under consideration.

Conditioning on previously-incorporated evidence is critical for inference—the imperative emerges directly from the mathematics of probability theory. However, reasoning about logical dependence in qualitative research can be extremely challenging. Evidence can be connected in arbitrarily complex ways; E_x and E_y may even be dependent in multiple ways, some of which might lead us to raise $P(E_x \mid H_j\ E_y\ I)$ relative to $P(E_x \mid H_j\ I)$, whereas others might lead us to lower $P(E_x \mid H_j\ E_y\ I)$ relative to $P(E_x \mid H_j\ I)$. Moreover, evidence may be dependent under some hypotheses but independent under others. In practice, we should aim to do the best we can while recognizing that it may not be feasible to carefully handle multiple nuanced dependencies in the data.

3.3.4. Sequence the Evidence as Convenient

The rules of conditional probability imply that the order of the evidence does not affect the final posterior probabilities. Using the product rule of probability and commutativity, the joint likelihood of two pieces of evidence can be written in any of the following equivalent ways:

$$P(E_A\, E_B|H\ I) = P(E_B\, E_A|H\ I) = P(E_A|E_B\ H\ I)\ P(E_B|H\ I) = P(E_B|E_A\ H\ I)\ P(E_A|H\ I). \quad (3)$$

Some literature in the psychological Bayesian tradition introduces non-standard rules for updating probabilities, implying that the order in which evidence is analyzed does matter (e.g.

[10] E.g. Beach and Pedersen (2013:128), Bennett and Checkel (2015:27-28).

Jeffrey 1983). However, such approaches necessarily violate the fundamental rules of probability (equation 3) and conflict with the notion of rationality that lies at the heart of logical Bayesianism—information incorporated in logically equivalent ways should lead to the same conclusions. If in practice we do not arrive at the same posterior probabilities upon analyzing the evidence in a different order, we have made a mistake somewhere in our reasoning that needs to be corrected.

Because we are free to consider the evidence in any order, we can look for sequences that facilitate conditioning on previously incorporated evidence when assessing likelihoods. Placing strongly discriminating evidence last can preclude having to condition other pieces of evidence on the conjunction of a hypothesis and an observation that is extremely implausible under that hypothesis—a difficult mental exercise that requires imagining very surprising flukes or coincidences. Incorporating highly-decisive evidence last can also obviate careful conditioning on previous evidence, because the likelihood ratios will be extremely large regardless. On the other hand, if the evidence is decisive enough, we could incorporate it first and be done, since additional evidence will contribute only marginally to our posteriors.[11]

3.4 Logarithmic Scales for Probabilities

Whereas most research on Bayesian process tracing assigns values to probabilities on essentially linear, often course-grained scales (e.g. Rohlfing 2013:19), we advocate a logarithmic scale for odds and likelihood ratios, which is standard practice in the natural sciences and information sciences. Using a logarithmic scale, in conjunction with an analogy to sound, better leverages intuition, enhances consistency when working with qualitative information, and may facilitate intersubjective agreement.

This recommendation is grounded in psychophysics, which shows that sensory perception tends to be a logarithmic function of stimulus strength. Stated in differential terms, a just-noticeable difference in the loudness of sound, brightness of light, or pressure on the skin is proportional to the magnitude of the stimulus. This relationship (the Weber-Fechner Law), which works well for a wide variety of stimuli and a large range of magnitudes, is sensible given that humans experience stimuli of highly varied intensity. By building in a logarithmic scale, evolution has increased the nervous system's dynamic range. Given this characteristic feature of the nervous system, a logarithmic scale is more natural than a linear scale for measuring and analyzing sensory inputs. Sound, for example, is measured in decibels (dB), such that increasing the intensity of the sound wave by a factor of ten corresponds to an additive increment of 10 dB.

For similar reasons, logarithmic scales were introduced to assess perceptions of uncertainty in probabilistic inference. Good's (1985) *weight of evidence* in favor of one hypothesis compared to a rival, measured in decibels, is proportional to the logarithm of the likelihood ratio:

$$WOE\,(H_i : H_j) = 10 \log_{10} [\, P(E|H_i\ I) / P(E|H_j\ I)\,] \qquad (4)$$

The weight of evidence describes the probative value of the evidence—how strongly it discriminates between two rival hypotheses. Good (1985) contends that a change in weight of evidence of one decibel, for example from even odds of 1:1 to odds of around 5:4, is as fine-grained as we can reliably quantify our degree of belief in competing hypotheses. A change in probability from 75% to 90% corresponds to an increase in log-odds of about 5 dB, which is salient, but in the natural sciences, cogent evidence might regularly lead to swings of several tens of decibels. Notice that in Bennett's (2015:281) illustration of a smoking-gun test, where $P(E|H)$

[11] See Appendix A for further exposition of these considerations as applied to our case study.

=0.2 and $P(E|{\sim}H)$ =0.05, the weight of evidence is only 6 dB—salient, but not decisive enough by Good's standards to serve as a smoking gun for H.

Measuring log-odds in decibels lets us leverage everyday experience with sound, while providing a quantitative underpinning for Gull's metaphor of Bayesian inference as a "dialogue with the data"—in essence we can ask whether the evidence is whispering or shouting in favor of a given hypothesis. In acoustics, the minimal noticeable change is roughly 3 dB. A change of 5 dB is clearly noticeable, while an increase of 10 dB is perceived as about twice as loud; 20 dB is roughly four times louder. Table 1 provides typical reference sounds in decibels.

In qualitative research, we suggest regarding decisive evidence that strongly favors one hypothesis over a rival as roughly 30 dB, the difference between a quiet bedroom and a conversation—in other words, the data are "talking clearly." Likewise, a very low prior log-odds against a hypothesis relative to a more plausible rival could reasonably be set at –50 dB (Jaynes 2003:99-100), the difference between a pin drop and a normal conversation.

Table 1: Typical Sound Levels (dB)

Reference scales vary across sources. See for example: www.osha.gov/dts/osta/otm/new_noise/

10	Adult hearing threshold; rustling leaves, pin-drop
20-25	Whisper
30	Quiet bedroom or library, ticking watch
45	Sufficient to wake a sleeping person
50	Moderate rainstorm
60	Typical conversation
70	Noisy restaurant, common TV level
80	Busy curbside, alarm clock
90	Passing diesel truck or motorcycle
100	Dance club, construction cite
115	Rock concert, baby screaming

3.5 Inference via Bayes' Rule

The last step in explicit Bayesian process tracing entails applying Bayes' rule to draw an inference. When working with multiple pieces of evidence, we can successively apply Bayes' rule (equation 1), or we can multiply likelihood ratios to conduct inference in a single step:

$$\frac{P(H_i|E\ I)}{P(H_j|E\ I)} = \frac{P(H_i|\ I)}{P(H_j|\ I)} \times \frac{P(E_1|H_i\ I)}{P(E_1|H_j\ I)} \times \frac{P(E_2|E_1\ H_i\ I)}{P(E_2|E_1\ H_j\ I)} \times \ldots \times \frac{P(E_N|E_1 \ldots E_{N-1}\ H_i\ I)}{P(E_N|E_1 \ldots E_{N-1}\ H_j\ I)}\ , \tag{5}$$

which follows from equations (1) and (2).

Taking the logarithm of the equation (5) gives a particularly simple, additive form of Bayes' rule—the posterior log-odds equals the prior log-odds plus the weight of evidence:

$$10 \log_{10} \left[\frac{P(H_k|E\ I)}{P(H_l|E\ I)}\right] = 10 \log_{10} \left[\frac{P(H_k|I)}{P(H_l|I)}\right] + WOE(H_k : H_l\ )\ . \tag{6}$$

This formulation offers the computation advantage that weights of evidence are also additive:

$$WOE(H_k : H_l) = WOE_1(H_k : H_l) + WOE_2(H_k : H_l, E_1) \ldots + WOE_N(H_k : H_l, E_1 \ldots E_{N-1}) \;, \qquad (7)$$

where we must remember to condition on E_{prev} as appropriate.

4. Empirical Example: Evaluating Weights of Evidence

Appendix A provides an explicit Bayesian analysis of a case study from Fairfield's (2013, 2015) research on tax policy change in Latin America. We compare Fairfield's hypothesis against three rivals in light of six key observations from the case narrative. We assign priors drawing on our background information, evaluate likelihoods for each piece of evidence, and use Bayes' rule to derive posterior probabilities. We then conduct Bayesian sensitivity analysis to ascertain how much our conclusions depend on choices of priors and values of likelihood ratios.

This section overviews our analysis of weights of evidence—the most important step in Bayesian inference—for two pieces of case evidence. Assessing weights of evidence, which involves ratios of likelihoods under alternative hypotheses, lets us ignore details that have a common effect on likelihoods irrespective of which hypothesis is true. For instance, we might judge the amount of effort expended to obtain an interview with an informant or to gain access to an archive to be largely independent of which hypothesis is correct. Similarly, this approach minimizes effort required for conditioning on previous evidence when an argument can be made that dependence does not differ much across hypotheses. Working directly with ratios also helps leverage intuition—while we might reason that E_x makes us ten times (10 dB) more confident in H_1 vs. H_2, we may be less comfortable assigning value to each of the respective likelihoods.

Fairfield's case study examines a Chilean reform that revoked a regressive tax subsidy. She argues that an "equity appeal," made during a presidential race in which inequality had assumed unusually high salience, compelled the right-wing opposition coalition to accept the reform in order to avoid electoral punishment. During the 2005 campaign, the right's presidential candidate blamed Chile's persistent inequality on the governing center-left coalition; the incumbent president responded by linking the tax subsidy to inequality and publicly challenging the opposition to support the reform. We denote this equity-appeal hypothesis as H_{EA}. Below, we consider just two rival hypotheses: H_I = *The opposition accepted the reform because Chile's institutionalized party system motivates cross-partisan cooperation and consensual politics*, and H_{CC} = *The opposition accepted the reform because its core constituency—business and upper-income individuals—had a weaker material interest in defending the tax subsidy in 2005 compared to previous years, due to a decline in the assets eligible for the subsidy*. We assume as part of our background information that H_{EA}, H_I, and H_{CC} are mutually exclusive and exhaustive.

E_1 *=Governing-coalition informants told the investigator that the center-left coalition discussed including a measure to eliminate the tax subsidy in multiple prior tax reforms, but that measure was ruled out as infeasible on every such occasion due to resistance from the right coalition.*

$WOE_1(H_{EA} : H_I) = 30$ dB. We judge E_1 to be speaking clearly in favor of H_{EA} relative to H_I—we would be far more surprised to observe E_1 in a world where H_I is the correct hypothesis. In the world of H_{EA}, the equity appeal altered the right's behavior from resisting the reform in previous years to accepting the reform in 2005. E_1 is consistent with this scenario. In contrast, if institutions produced consensus on eliminating the tax subsidy in 2005, they should also have done so in previous years, since our background information includes the fact that institutions did not change during the intervening period. If eliminating the tax benefit had been discussed and ruled out on only one occasion, we would be less surprised; however, E_1 is a conjunction of multiple instances in which institutions failed to promote right-party cooperation.

WOE_1 (H_{EA} : H_{CC}) = 3 dB. While E_1 is consistent with both H_{EA} and H_{CC}, we view the weight of evidence as weakly favoring H_{EA}. In the world of H_{CC}, the right accepted the reform in 2005 because its core constituency no longer had sizable assets that benefitted from the tax subsidy, even though our background information (I) indicates that the right tended to resist even modest tax increases as a matter of principle. To avoid a contradiction between H_{CC} and I, we must assume that by 2005, the asset decline had pushed the right past its threshold of resolve for resisting the reform. However, under H_{CC} and I, we have no clear prediction about when exactly we would expect the tax subsidy to be eliminated, whereas under H_{EA} and I, we have a clear rationale for why the reform occurred in 2005 as opposed to some years earlier or later. Under H_{CC}, the observed timeline is just one of several more or less equally plausible scenarios in which the successful reform occurs in a different year. These multiple possibilities reduce the likelihood of the specific timeline observed.

E_1 illustrates the importance of comparing specific alternative hypotheses rather than attempting to evaluate the working hypothesis directly against its full logical negation. Section 3.1 emphasized that assessing likelihoods conditional on an ill-specified $\sim H$ may be essentially impossible. For E_1, the weights of evidence in favor of H_{EA} relative to the two rivals are not equal. If we try to directly evaluate WOE_1 (H_{EA} : $\sim H_{EA}$), how would we mentally inhabit a world as vague as $\sim H_{EA}$, which is really an amalgamation of two very different kinds of worlds? Only after we have decomposed $\sim H_{EA}$ into H_I and H_{CC} can we reason meaningfully about weights of evidence by inhabiting each alternative world in turn.

E_2* = *A finance ministry official told the investigator that the tax subsidy "was a pure transfer of resources to rich people... It was not possible for the right [coalition] to oppose the reform after making that argument about inequality." Likewise, the former President told the investigator that the tax subsidy "never would have been eliminated if I had not taken [the opposition candidate] at his word" when the latter publicly professed concern over inequality.

We have aggregated two interview excerpts into a single piece of evidence, because they are strongly dependent under any hypothesis. Regardless of which explanation is correct, we expect that the president and finance ministry officials have communicated extensively and share similar analyses of why the right accepted the reform. Our aggregation decision also follows the rule of thumb from §3.2; we have similar information from similar sources.

WOE_2 (H_{EA} : H_I, E_1) = WOE_2 (H_{EA} : H_{CC}, E_1) = 10 dB. E_2 provides glimpses of the causal mechanism underlying H_{EA} by referring to the exchange between the opposition candidate and the president that culminated in the equity appeal. Because E_2 makes the government appear savvy and effective at achieving socially-desirable goals while highlighting the opposition's resistance to redistribution, we see little reason for the government to conceal this information if H_{EA} is true. The likelihood $P(E_2 \mid H_{EA}\, E_1)$ should therefore be high. In an alternative world where either of the rival hypotheses is true, government informants might nevertheless have incentives to attribute the opposition's support for the 2005 reform to the equity appeal, because this story portrays the government in a positive light and the opposition in a negative light. However, our background information gives us confidence in these informants' knowledgeability, analytical judgments, and sincerity. Balancing these considerations, we judge E_2 to favor H_{EA} over each rival by 10 dB.

E_2 illustrates that the accuracy of the information provided by the sources depends on the hypothesis (§3.3.2). Under H_{EA}, the informants' statements must be taken as true, whereas under the rivals, the statements are necessarily false—the informants are either mistaken or lying.

5. Improving Inference and Analytic Transparency

In the context of efforts to establish process tracing as a rigorous methodology and growing attention to analytical transparency, several prominent scholars have argued that explicitly employing Bayesian analysis will make inferences more systematic and more amenable to scrutiny (Rohlfing 2013; Bennett 2015:297; Humphreys and Jacobs 2015). We concur that explicit Bayesian analysis, if carefully implemented, offers several important advantages and opportunities for improving inference. This approach forces us to clearly identify and carefully consider all salient evidence as well as critical elements of our background information. It precludes subconsciously focusing on a favored working hypothesis by requiring us to consider states of the world characterized by rival hypotheses. Explicit Bayesian analysis may also "eliminate the considerable ambiguity in many verbal phrases used to convey probabilities" (Bennett 2015:297); in particular, the decibel scale we advocate could help communicate our degrees of belief more effectively. With practice, the guidelines elaborated in Section 3 may foster more systematic and more logically rigorous inferences compared to untrained intuition.

However, explicit Bayesian analysis is clearly a "very tall order" (Humphreys and Jacobs 2015:42), especially for evidence-intensive process tracing. As such, we must weight the advantages against the potential limitations and drawbacks (Table 2). We begin by discussing caveats based on our experience of elaborating Appendix A and then assess when explicit Bayesian process tracing will be more valuable or less valuable in light of the challenges.

Table 2: Assessing the Potential of Explicit Bayesian Analysis

Advantages and opportunities	Caveats and limitations
• Ensures clear identification and careful assessment of salient evidence and key elements of the background information • Ensures consideration of alternative hypotheses • Facilitates more effective communication of degrees of belief • May help inferences approximate Bayesian logic more closely than untrained intuition	• Probabilities cannot be unambiguously specified • Requires substantial training • Intractable beyond relatively simple causal models • May be unduly burdensome in practice • Explicit Bayesian analysis and transparency are not synonymous

5.1 Caveats and Limitations

The foremost challenge of explicit Bayesian process tracing entails assigning numerical values to all probabilities (priors and likelihoods). In the natural sciences, a strong underlying theory and a description of the measurement apparatus (along with various simplifying assumptions) lead to a specific error model and thereby a likelihood function that unambiguously specifies probabilities for the various possible measurements that may be observed. In the social sciences, there is no clear procedure for translating complex, narrative-based, non-reproducible, often qualitative information into precise probability statements. Efforts to explicitly apply Bayesian learning in qualitative social science will inevitably run up against the difficulty of quantifying likelihoods without precision or objectivity. Specifying a range of probabilities rather than a precise value (Humphreys and Jacobs 2015) may be adequate for some purposes, but ultimately, it simply relocates the arbitrariness of assigning numerical values.[12] Even from a purely pragmatic perspective, this sort of approach has limited value, because propagating interval probabilities through Bayes' rule is non-trivial when working with multiple hypotheses. While words used to express probability in common parlance are certainly ambiguous, quantification may simply disguise that ambiguity with false precision.

A second caveat is that substantial training may be necessary before explicit Bayesian analysis can improve upon intuition. Our teaching experience thus far indicates that a day or two of intensive workshops is not adequate to successfully apply this approach. In fact, as part of the learning process, reasoning may get worse before getting better. While the metaphor of weighing evidence is in some ways intuitive, assessing the weight of evidence changes the way that we use our intuition. Practice will be needed to master this fundamental aspect of Bayesian thinking.

Third, explicit Bayesian analysis becomes intractable beyond fairly simple causal models, which are rarely adequate in social science. Recall that Bayesian analysis entails specifying mutually-exclusive hypotheses, which is nontrivial and may require over-simplification. Some of the hypotheses we assess against Fairfield's (2013) explanation involve causal mechanisms that—in the real world—could potentially operate simultaneously or in interaction. Assessing such possibilities requires elaborating a more complex but still mutually-exclusive hypothesis space, which would aggravate the challenges of assigning numerical values to likelihoods. By contrast, in the natural sciences, Bayesian analysis is usually applied to very simple hypothesis spaces (even if the underlying theory and experiments are highly complex); for example: H_1 = the Higgs boson mass is 124–126 GeV/c^2, H_2 = the mass is 126–128 GeV/c^2, etc.

Fourth, practical considerations may restrict the use of explicit Bayesian analysis. Appendix A, which includes various cross checks to ensure that likelihoods are sensible and reasonably consistent across the problem, exceeds the length of Fairfield's (2013) article, which included three additional case studies; Fairfield's (2015) book includes over 28 case studies. Assessing weights of evidence (likelihood ratios) instead of absolute likelihoods (Section 4) is a more feasible if somewhat less careful approach. But either way, explicit Bayesian analysis cannot replace case narratives, which are an effective and efficient way to holistically communicate context, evidence, and analysis. As such, *requiring* scholars to provide explicit Bayesian analysis for all of their cases, above and beyond their narrative work, would create heavy disincentives for process tracing.

Finally, explicit Bayesian analysis should not be directly equated with transparency. On the one hand, this approach can obscure rather than clarify inference, especially if we disaggregate the evidence too finely and unpack our analysis into too many steps—we may become lost in minutiae. Moreover, making too many steps explicit may lull readers into uncritically accepting or glossing over the author's reasoning, rather than assessing whether they can arrive at the conclusions through their own independent logical pathways, thereby undermining the scholarly scrutiny of inferences that analytical transparency is intended to promote. Even mathematicians routinely skip steps in published proofs; readers must fill in and verify themselves, which provides an important cross-check. On the other hand, transparency does not require assigning numerical values to probabilities and applying Bayes' rule to derive an inference. Scholars can make the assumptions and logic behind their inferences explicit without resorting to numbers.

[12] To avoid subjective likelihood assignments, Humphreys and Jacobs (2015) include priors on the probative value of process-tracing clues; yet the problem then becomes how to translate background knowledge and theoretical expectations into an appropriate prior distribution. Moreover, if we work within a single case, only averages over priors for clue probabilities matter, so their approach reduces to specifying likelihoods.

While these considerations do not necessarily constitute an argument *against* explicit Bayesian process tracing, they clarify that transparency is not necessarily an argument *for* adopting this approach.

5.2 Applications

Given the caveats, when might explicit Bayesian analysis prove most useful? We begin with situations where we expect this approach to be of limited value and proceed toward those that proffer higher value. If all observations strongly favor a particular hypothesis over the rivals, explicit Bayesian analysis is unlikely to improve on intuition. Scholars can explain why the evidence is decisive without quantifying probabilities, and if the evidence is indeed compelling, readers should recognize it as such on its face. Likewise, if the evidence has weak probative value, explicit Bayesian analysis may simply confirm the realization we would have obtained intuitively—the evidence is insufficient to strongly support any particular hypothesis (unless we already had strong priors or cannot think of reasonable alternatives).

Moderate gains for inference may arise when the evidence is complex and does not clearly favor a single hypothesis. On the one hand, when some observations favor one hypothesis whereas others favor a rival, it may be difficult to intuitively ascertain which provides the best explanation. Explicit Bayesian analysis helps us keep track of nuances, consistently assess the weight of evidence for each observation, and systematically aggregate inferences across observations. On the other hand, there is a danger when evidence is ambivalent that conclusions derived via explicit Bayesian analysis may be driven by arbitrariness inherent in assigning numerical values to probabilities. Natural scientists would only believe that noisy data accumulates into a significant signal if the error model is well understood; in qualitative social science, analogous situations may rarely arise. Almost by definition, if the evidence pulls in different directions, small changes in probabilities may swing the inference in favor of one hypothesis or another.

Nevertheless, explicit Bayesian process tracing in such cases may be merited for the sake of analytical transparency and informing future research. Regarding transparency, if we must make inferential claims on the basis of ambivalent or weak evidence—if important questions are at stake[13] and obtaining better evidence is infeasible—explicit Bayesian analysis can at least clarify the basis on which those claims rest and facilitate debate among scholars. Looking forward to future data-gathering opportunities, explicit Bayesian analysis could also elucidate what kinds of additional evidence would be most useful for strengthening the inference.

We envision a more valuable role for explicit Bayesian analysis in identifying the locus of contention when scholars disagree on inferences. As Hunter (1984:88) argues, through Bayesian analysis, "the sources of the disagreement can be determined much more easily than in normal verbal analysis." Explicit Bayesian analysis provides a clear framework for pinpointing disagreements: Do they arise from different background information and assumptions (e.g. a source's motives or sincerity), different priors, or different assessments of likelihoods? If the problem lies with the probative value of evidence, which observations are most contested, and why? For these purposes, numbers serve primarily to stimulate discussion about inferential logic, assumptions, and judgments, and the ad-hoc component of quantification will be less problematic. Our Bayesian sensitivity analysis in Appendix A illustrates how this clarification and adjudication process might work; we show that to remain unconvinced, a skeptical reader

[13] Hunter (1984) explores military applications.

would need to have extremely strong priors against the equity appeal explanation and/or contend that the evidence is far less discriminating than we have argued.

We also foresee a valuable pedagogical role for explicit Bayesian analysis. Reading examples and conducting exercises can train intuition to follow this inferential logic more systematically, thereby improving traditional narrative-based process tracing. For example, one of the most salient lessons from our empirical case application is that the weight of evidence depends by definition on which hypotheses we compare; we cannot judge how decisive the evidence is with respect to our working hypothesis alone, without considering concrete alternatives. Thinking in terms of the weight of evidence, even without assigning numbers, may help scholars identify their most discriminating observations. More generally, many of our guidelines for explicit Bayesian analysis (summarized in Table 3) have direct analogs for more heuristic Bayesian reasoning in narrative-based process tracing—from comparing clearly-specified mutually exclusive hypotheses and explaining why our background information suggests some explanations and justifies disregarding others from the outset, to considering how a source's potential biases and instrumental incentives might change under rival hypotheses, and thinking about logical dependencies among the evidence when drawing inferences.

Relatedly, elaborating an explicit Bayesian appendix for an illustrative case from one's own research might help establish process-tracing "credentials." As much as we try to make our analysis transparent, multiple analytical steps will inevitably remain implicit. Qualitative research draws on vast amounts of data, often accumulated over multiple years of fieldwork. There is simply too much evidence and too much background information that informs how evidence is evaluated to fully articulate or catalog. Qualitative research is not replicable as per a controlled laboratory experiment; at some level, we must trust that scholars have made sound judgments. To that end, scholars might use an explicit Bayesian illustration to demonstrate their care in reasoning about the evidence and the inferences it permits.

Table 3: Guidelines for Explicit Bayesian Process Tracing

Hypotheses and Priors	Articulate clearly-specified, mutually exclusive hypotheses; do not attempt to directly compare *H* vs. ~*H* Choose an option for assigning prior probabilities: a) Specify priors that aim to reflect background information as best as possible b) Assign equal priors following the "principle of indifference" c) Conduct the analysis using multiple different prior distributions
Evidence	Treat information from distinct types of sources as separate pieces of evidence To facilitate reasoning, seek a middle ground between aggregating evidence into overly course-grained pieces and disaggregating into excessively fine-grained pieces
Likelihoods $P(E \mid H_x\, E_{prev}\, I)$	Mentally "inhabit" the world of each hypothesis Assess the likelihood that "source S stated X", considering that biases and instrumental incentives may change under rival hypotheses Condition likelihoods on all previously-incorporated pieces of evidence, thinking carefully about logical dependence Sequence the evidence as convenient to facilitate reasoning about logical dependence Identify key elements of the background information and explain how they inform likelihoods Simplify analysis by evaluating likelihood *ratios* and/or weight of evidence
Logarithmic Scales	Assign numerical values to probabilities on a logarithmic scale Use an analogy to sound decibels to enhance consistency and leverage intuition

References

Abell, Peter. 2009. "A Case for Cases: Comparative Narratives in Sociological Research." *Sociological Methods and Research* 38:38-70.

Barrenechea, Rodrigo, and James Mahoney. 2016. "A Set-Theoretic Approach to Bayesian Process Tracing." Syracuse Institute for Qualitative and Multi-Method Research.

Beach, Derek, and Rasmus Pedersen. 2013. *Process-Tracing Methods.* University of Michigan.

Bennett, Andrew. 2008. "Process Tracing: A Bayesian Perspective." In Janet Box-Steffensmeier, Henry Brady, and David Collier, eds, *The Oxford Handbook of Political Methodology*, Oxford University Press, 702-721.

_____. 2015. "Disciplining Our Conjectures: Systematizing Process Tracing with Bayesian Analysis." In Andrew Bennett and Jeffrey Checkel, eds, *Process Tracing in the Social Sciences: From Metaphor to Analytic Tool*. Cambridge University Press, 276–98.

Bennett, Andrew, and Jeffrey Checkel, eds. 2015. *Process Tracing in the Social Sciences: From Metaphor to Analytic Tool*. Cambridge University Press.

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

Cox, Richard. 1961. *The Algebra of Probable Inference.* Johns Hopkins University Press.

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

———. 2015. *Private Wealth and Public Revenue in Latin America: Business Power and Tax Politics.* Cambridge University Press.
Gelman, Andrew, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. 2013. *Bayesian Data Analysis*, *Third Edition.* CRC Press.
Gill, Jeff. 2008. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Taylor & Francis.
Good, I.J. 1985. "Weight of Evidence: A Brief Survey." In J.M. Bernardo, M.H. de Groot, D.V. Lindley, and A.F.M. Smith, eds., *Bayesian Statistics 2.* New York: Elsevier.
Gregory, Phil. 2005. *Bayesian Logical Data Analysis for the Physical Science.* Cambridge.
Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach.* Caris Publishing Company.
Hunter, Douglas. 1984. *Political/Military Applications of Bayesian Analysis*. Boulder: Westview.
Humphreys, Macartan, and Alan Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review.* 109 (4):653-73.
Jackman, Simon. 2009. *Bayesian Analysis for the Social Science.* Wiley.
Jaynes, E.T. 2003. *Probability Theory: The Logic of Science.* Cambridge University Press.
Jeffery, Richard. 1983. *The Logic of Decision.* University of Chicago Press.
LaPlace, Pierre Simon. 1812. *Analytical Theory of Probabilities*.
Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods and Research* 41: 570-97.
McKeown, Timothy. 1999. "Case Studies and the Statistical Worldview." *International Organization* 53 (1): 161-190.
QMMR 2015. Symposium on Qualitative and Multi-Method Research. APSA Quantitative and Multi-Method Research Section Newsletter, Vol 14 (1).
Rohlfing, Ingo. 2013. *Case Studies and Causal Inference.* Palgrave Macmillan.
Savage, Leonard. 2003. *The Foundations of Statistics.* New York: Wiley.
Sivia, D.S., 2006. "Data Analysis—A Dialogue With The Data," in Advanced Mathematical and Computational Tools in Metrology VII, P. Ciarlini, E. Filipe, A.B. Forbes, F. Pavese, C. Perruchet and B. Siebert (eds.), World Scientific Publishing Co., 108-118.
Stokes, Susan. 2001. *Mandates and Democracy: Neoliberalism by Surprise in Latin America.* Cambridge University Press.
Tannenwald, Nina. 2007. *The Nuclear Taboo.* Cambridge University Press.
Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Cornell.