Updating Bayesian(s): A Critical Evaluation of Bayesian Process Tracing

Sherry Zaks [ID]

University of Southern California, Department of Political Science, 3518 Trousdale Pkwy, VKC 327, Los Angeles, CA 90087, USA.
Email: szaks@usc.edu

Abstract

Given the increasing quantity and impressive placement of work on Bayesian process tracing, this approach has quickly become a frontier of qualitative research methods. Moreover, it has dominated the process-tracing modules at the Institute for Qualitative and Multi-Method Research (IQMR) and the American Political Science Association (APSA) meetings for over five years, rendering its impact even greater. Proponents of qualitative Bayesianism make a series of strong claims about its contributions and scope of inferential validity. Four claims stand out: (1) it enables causal inference from iterative research, (2) the sequence in which we evaluate evidence is irrelevant to inference, (3) it enables scholars to fully engage rival explanations, and (4) it prevents ad hoc hypothesizing and confirmation bias. Notwithstanding the stakes of these claims and breadth of traction this method has received, no one has systematically evaluated the promises, trade-offs, and limitations that accompany Bayesian process tracing. This article evaluates the extent to which the method lives up to the mission. Despite offering a useful framework for conducting iterative research, the current state of the method introduces more bias than it corrects for on numerous dimensions. The article concludes with an examination of the opportunity costs of learning Bayesian process tracing and a set of recommendations about how to push the field forward.

Keywords: Bayesian methods, qualitative methods, process tracing, Bayesian analysis

Qualitative research is plagued by two unresolved debates. First is the unresolved question of what process tracing *is* in the first place—whether analytic narratives suffice or we should use evidentiary tests, whether it is rooted in formal logic or Bayesian logic, and the debates go on.[1] Second is the debate over what scholars need to do for their qualitative work to be deemed rigorous—and whether *any* approach can be trusted without submitting to the rigid transparency guidelines spearheaded by the Data Access and Research Transparency (DA-RT) initiative.[2] The ontological, epistemological, and ethical cleavages of these disputes run wide and deep. With scholars boycotting journals, and—effectively—journals boycotting scholars over transparency issues, the stakes of resolving these debates are high. As such, any approach that aims to resolve both issues within a unified framework deserves close attention. Proponents of Bayesian process tracing claim to do just that.

The Bayesian approach provides an analytic template for asking (and answering) a central research question: *What is the probability that our main hypothesis (H_M) is correct, given that we searched for and found evidence E_i?* Using Bayes' rule, researchers are encouraged to answer this question by explicitly specifying quantities such as the prior probability their hypothesis is correct given only their background information, and the likelihood of observing each piece of evidence in a world where their main hypothesis is true. Together, these quantities allow researchers to

---

1   On analytic narratives, see George and McKeown (1985). On evidentiary tests, see Van Evera (1997); Brady, Collier, and Seawright (2010). On formal logic approaches, see Mahoney (2012); Goertz and Mahoney (2012). On Bayesian process tracing, see Bennett (2009); Beach and Pedersen (2012, 2019); Bennett (2014); Bennett and Checkel (2015); and Fairfield and Charman (2017, 2019).

2   Moravcsik (2014); Lieberman (2016); Yom (2018). For further reading into this debate, see the qualitative transparency deliberations housed at https://www.qualtd.net.

update their confidence in a given theory as they move through the evidence. Proponents argue that explicitly specifying probabilities for the quantities in Bayes' rule not only makes process tracing more transparent,[3] but for some, it makes process tracing sufficiently transparent as to obviate some of the burdensome recommendations of the DA-RT initiative.[4]

While the tradition of using Bayesian inference as a metaphorical frame for qualitative work dates back nearly forty years,[5] a recent wave of work in this vein has laid out a set of research practices and a corresponding set of strong claims about what explicit Bayesianism (i.e., specifying the probabilities in Bayes' rule and updating mathematically) is capable of achieving.[6] Beyond asserting its foundational role in process tracing, four claims stand out: (1) the Bayesian approach enables causal inference from iterative research, (2) Bayesian logic makes the sequence in which we evaluate evidence irrelevant to inference, (3) the approach enables scholars to more fully engage rival explanations, and (4) Bayesian logic guards against ad hoc hypothesizing and confirmation bias.[7] If these assertions hold, the Bayesian approach will constitute a revolutionary advancement in process tracing.

Given the increasing quantity and impressive placement of work in this vein, Bayesian approaches represent a new frontier of qualitative research methods. Moreover, they have made a substantial impact in the discipline beyond the press. Since 2014, the Process Tracing modules at the Institute for Qualitative and Multi-Method Research (IQMR) and the American Political Science Association (APSA) Short Courses have focused primarily on the Bayesian approach, which means that for many students, this approach is the first and last word they receive in the way of qualitative training. Additionally, the memorandum put forth by the process-tracing subcommittee of the Qualitative Transparency Deliberations frames process tracing as though it is primarily a Bayesian method.[8] Finally, Bennett argues that the Bayesian approach can be adopted to policy analysis to improve projections and decision-making.[9] As yet, however, no one has conducted a systematic evaluation of the promises, trade-offs, and limitations of the Bayesian approach in practice. In light of both its growing footprint and ambitious claims, I take a step back to critically evaluate whether and to what extent the method lives up to the mission—and what happens when it comes up short.

This article proceeds as follows. I begin with an overview of Bayesian process tracing. The following four sections correspond to the claims outlined above. I lay out the stakes of each claim and evaluate the extent to which Bayesianism is the best tool for the job. I demonstrate that Bayesian tools add value to some areas of the research process, but they come with severe and often unacknowledged shortcomings that demand resolution before this approach can be widely adopted. I conclude by evaluating what I call the implicit fifth claim: that the analytic benefits of investing in the Bayesian approach outweigh the opportunity costs of dedicating methodological training time to other endeavors. Finally, I enumerate potential avenues for future development.

1 Overview of Qualitative Bayesian Research

This section gives an overview of Bayesian inference to contextualize the discussion. As Fairfield and Charman succinctly note, "Bayesian reasoning is simply a process of updating our views about which hypothesis best explains the…outcome of interest as we learn additional information."[10]

---

3 Abell (2009, 61); Beach and Pedersen (2012, 83); Bennett (2015a, 297).
4 Fairfield and Charman (2019).
5 See Putnam (1981, loc. 3623) and McKeown (1999).
6 Abell (2009); Bennett (2009, 2014, 2015a); Beach and Pedersen (2012, 2019); Humphreys and Jacobs (2015); Fairfield and Charman (2017, 2019).
7 These claims are most explicitly laid out in Fairfield and Charman (2019), but most are echoed throughout the Bayesian process-tracing literature.
8 Indeed, a simple text search reveals that (save the references) the word "Bayesian" appears more than the phrase "process tracing" Bennett, Fairfield, and Soifer (2019).
9 Bennett (2015b, 228).
10 (2019, 158).

To anyone new to Bayesian reasoning, defining the process of continuous updating in light of evidence may seem so obvious as to not need a name beyond *research*. For proponents of Bayesian methods, the obviousness is a feature, not a bug, since a core goal of the Bayesian approach is to formalize the approach researchers use intuitively.[11] Given the central role of examining alternative hypotheses in addition to our own, Bennett as well as Fairfield and Charman advocate using the odds-ratio form of Bayes' rule, which assesses the strength of one hypothesis (H_M) relative to an alternative (H_A):

$$\underbrace{\frac{P(H_M|E_i I)}{P(H_A|E_i I)}}_{\text{Posterior}} = \underbrace{\frac{P(H_M|I)}{P(H_A|I)}}_{\text{Prior}} \times \underbrace{\frac{P(E_i|H_M I)}{P(E_i|H_A I)}}_{\text{Likelihood}}. \tag{1}$$

Beginning with the right-hand side of the equation, the numerator, $P(H_M|I)$, represents the *prior* probability on our main hypothesis: our degree of belief that H_M describes the world given only our background information, I. The denominator represents our prior on the alternative hypothesis that we are testing against the main.

The next term in the equation represents the *likelihood ratio*: the probability of observing a given piece of evidence in a world where H_M is true over the probability of observing the same piece of evidence in a world where H_A is true. The goal of the likelihood ratio is to formalize our assessment of how likely we are to find a given piece of evidence under competing hypotheses.

Finally, the prior and likelihood ratios are multiplied together to compute the *posterior* probability. The posterior gives us our *updated* confidence in each hypothesis given that we have observed evidence E_i. If, for example, E_i is more plausible in the world of H_M than the world of H_A (as assessed in the likelihood ratio), then our confidence in H_M relative to H_A will increase. Posterior probabilities (or ratios) are then used to inform the priors for analyzing subsequent pieces of evidence.

While some qualitative methodologists encourage scholars to use the Bayesian framework primarily as a metaphorical tool for updating their confidence in hypotheses in light of new evidence,[12] others—and the work that has made the largest footprint in the discipline—encourage an explicit mathematical implementation of Bayes' rule.[13] In the latter approach, scholars are encouraged to conjure and justify numerical quantities for the prior and the likelihood to mathematically compute our updated confidence in one hypothesis relative to another.[14] Bennett neatly captures the consensus among qualitative Bayesians, arguing that "explicitly assigning priors and likelihood ratios and using Bayesian mathematics can make process tracing more rigorous and transparent."[15] Scholars in this camp argue that while the probabilities themselves might be incorrect, the act of explicitly justifying them and subjecting both the probabilities and justifications to scrutiny (via peer review) contributes to greater transparency of the assumptions researchers make implicitly (even without a Bayesian approach).[16]

Due in part to the strength of the claims in the explicit Bayesian literature, and in part to the impact this approach has made in the discipline—both in the press and in methodological training modules—the remainder of the article focuses primarily on the explicit (mathematical) Bayesian approach. Explicit Bayesian process tracing requires the most time dedicated to learning the method, yet, according to its proponents, the investment is justified because it is capable

---

11 McKeown (1999, 179); Bennett (2009, 708); Fairfield and Charman (2019, 157).
12 McKeown (1999); Beach and Pedersen (2019).
13 Bennett (2014, 2015a); Fairfield and Charman (2017, 2019).
14 For examples of the explicit mathematical approach in practice see Bennett (2015a, 279–285), Humphreys and Jacobs (2015, 658), and Appendix A in Fairfield and Charman (2017).
15 Bennett (2015a, 297).
16 Bennett (2015a, 297); Fairfield and Charman (2017, 378); (2019, 163).

of delivering the greatest returns in terms of rigor, transparency, and the quality and scope of inferences.

2   Claim 1: Bayesianism Allows us to Revalue and Implement Iterative Research

Bayesians note that the research process often involves a messy "dialogue with the data,"[17] yet, the print version more closely resembles a monologue that begins in an armchair and ends with a light bulb. By formalizing guidelines for iteratively updating our confidence in hypotheses in light of additional evidence, the Bayesian approach may foster a more systematic and transparent account of how our research proceeds in practice. However, the attempt to move beyond the metaphor of Bayesian updating raises three questions and corresponding issues that must be resolved before the method is widely adopted.

First, the Bayesian process-tracing literature exhibits an unacknowledged debate about what "iterative research" refers to. While all qualitative Bayesian scholars agree that one aspect of iterative research is the process of updating our confidence in theories as we move through evidence, some also take the view that iterative research entails updating the theories as well.[18] For example, where Bennett argues that Bayesianism is not suited to theory generation,[19] Beach and Pedersen contend that "in most realistic research situations, theory-building and theory-testing are actually used in an iterative fashion."[20] Thus, for some, Bayesian logic provides a template for the inductive side of research.

If one of the core goals of bringing explicit Bayesian computation to process tracing is to provide a formal template for the processes scholars tend to use implicitly, acknowledging that scholars often update theories as they analyze evidence works in service of this goal. However, if iteration does involve an inductive component, Bayesian methodologists must provide clear guidelines for devising and altering priors on hypotheses as the hypotheses themselves change. Additionally, the literature must instruct scholars on how to deal with evidence that has already been analyzed before the hypothesis was altered. Currently, however, the literature is silent on these crucial matters. Thus, for those who claim that Bayesian iteration works in service of theory refinement, there is more work to do before the method can deliver on this goal.

Second, current formulations of the procedure lack clear and consistent guidelines for implementing a systematic, iterative process in practice. The classical formulation of Bayes' theorem (which tests a hypothesis against its logical negation) enables researchers to update the probability that a single hypothesis is true given a piece of evidence.[21] It provides a clear path forward for iteration by using the updated posterior as the new prior on that hypothesis for analyzing the next piece of evidence, and so forth. Yet, this form of Bayes' theorem is problematic because of the assumptions underlying the equation[22] and because it does not adjudicate among rival hypotheses, which is a central goal of process tracing.[23]

Fairfield and Charman offer a corrective via the odds-ratio form of Bayes' rule (Equation (1)). While this equation allows for adjudication by returning an update on the relative odds of two competing hypotheses given a piece of evidence, the process of iteration becomes elusive. Instinctively, the next step should be to use the posterior odds as the updated prior to analyze the next piece of evidence, E_2, but in the example given, the authors instead revert to placing equal

---

17 Quote from Fairfield and Charman (2019, 154), but also see McKeown (1999, 180); Bennett (2014, 47); and Beach and Pedersen (2012, 2019) for similar depictions of the research process.

18 Fairfield and Charman argue that the "iterative process of analyzing data" is fundamentally about "revising and refining theory," which, in turn, breaks down the line between inductive and deductive research (2019, 156).

19 Bennett (2015a, 276).

20 Beach and Pedersen (2019, 269).

21 Beach and Pedersen (2012) and Bennett (2014, 2015a) use this equation.

22 Zaks (2017, 354).

23 Zaks (2017, 354); Fairfield and Charman (2019, 158).

odds on the two hypotheses despite having previously found support for only one of them.[24] This choice calls into question the purpose of computing the posterior and the claim that Bayesian logic provides a formal template for iterative research. Moreover, the problem of how to conduct systematic iterative research becomes orders of magnitude more complex when researchers have more than two competing hypotheses.[25] If the Bayesian method does not establish clear guidelines for updating our confidence in hypotheses as we move through evidence, its utility as a tool for systematic iterative research becomes questionable.

Third, this recommendation raises questions of how narratives of iterative testing and refinement are expected to fit within the standard word limits of academic journals. Current disciplinary conventions already place space constraints on qualitative scholars; it would be harder still to fit in all the hypotheses we *thought* might explain the case first. Though some rightfully note that scholars can sidestep rigid word limits with online appendices or supplementary material,[26] this recommendation does not acknowledge the additional burden placed on qualitative scholars to produce "article-length" manuscripts that may effectively be one and a half or two articles worth of material for each publication. Of course, if Bayesian process tracing proves superior to other methods on dimensions of analytic rigor, inferential nuance, or transparency, this extra work may be deemed worthwhile, but the value added must be clear and measurable before enjoining scholars to take on that burden.

To be sure, qualitative Bayesian scholars are correct in noting that the pressure for flawless causal identification, proof of exogeneity, and preregistration of research designs belie the process by which a lot of research unfolds. Moreover, structuring analytic narratives around an iterative process has additional benefits beyond what Bayesian proponents acknowledge. Specifically, by omitting reference to how our theories developed over the course of a project, we lose a codified record of the ideas that seemed good on paper, but failed to pan out in testing. The discipline would benefit from embracing a model of research that not only accounts for the evolution of the research process, but also prevents other researchers from traversing the same (often well-trod) path to a dead end.

Across the qualitative Bayesian literature, scholars converge on the core benefit of this approach: Bayesian logic provides an explicit procedure for iterative updating as researchers move through evidence.[27] Although Bayesian updating provides a useful metaphorical tool for conducting iterative research, current formulations of the procedure lack the necessary guidelines for executing a systematic, iterative process in practice. Thus, in its current state, Bayesian process tracing needs to be refined further to ensure that this method is the best path for revaluing and implementing iterative research.

3   Claim 2: Timing of Evidence is Irrelevant to Inference

Qualitative Bayesians' second core claim is that the order of evaluating evidence and "keeping track of what we knew when" is "logically irrelevant" to inference.[28] This assertion aims to break down what Bayesians view as the hard—yet artificial—line drawn between exploratory (hypothesis-generating) and confirmatory (hypothesis-testing) stages of research.[29] Based on the mathematical irrelevance of time in probability theory, Fairfield and Charman push further than others in the Bayesian camp and articulate three strong claims about the unimportance of

---

24 Fairfield and Charman (2019, 162).
25 Specifically, researchers face a combinatorics problem in which testing two hypotheses requires as many analyses as there are pieces of evidence. Testing three hypotheses requires three analyses for each piece of evidence, and testing four hypotheses requires twelve analyses per piece of evidence. See Appendix A in the supplementary material for an elaborate explanation of the iteration problem.
26 Yom (2018, 420).
27 Bennett (2009); Beach and Pedersen (2012, 2019); Bennett (2014); and Fairfield and Charman (2017, 2019).
28 Fairfield and Charman (2019, 160); Bennett (2015a, 293).
29 Fairfield and Charman (2019, 156); Beach and Pedersen (2019, 174).

timing across the research process, enumerated below. While not all qualitative Bayesians echo these arguments, the implications that follow affect the entire research process, and thus, are substantial enough to warrant special attention.

(1) "New evidence has no special status relative to old evidence,"[30]
(2) "Learning the same pieces of information in different orders must produce identical results,"[31]
(3) "There are few analytical benefits to reporting temporal details about how the research process unfolded."[32]

The stakes of these arguments are quite high. If, indeed, logical Bayesianism enables scholars to test an inductively inspired hypothesis with the evidence that inspired it, this method not only calls into question the utility of preregistration and time-stamping evidence, but moreover, it suggests a much wider scope of inferential validity than the discipline currently embraces. If, however, the method falls short in practice, it could lead scholars to draw hasty or biased conclusions based on insufficient evidence.

The claims about the irrelevance of timing have implications for every stage of the research process, which I examine in turn: (1) hypothesis generation, (2) hypothesis refinement, (3) hypothesis testing, and (4) how researchers report results. While some of the arguments raise important points, I demonstrate that the focus on timing's *mathematical* irrelevance will cause those who fully subscribe to them to hastily disregard the cognitive, inferential, and practical relevance of timing considerations in conducting and documenting research. The corresponding problems cascade through all stages of research and have deleterious effects on inference.

3.1 Hypothesis Generation

Beginning with hypothesis generation, I demonstrate that for timing to be truly irrelevant, research cannot be costly (which it is) and the distribution of support for a given hypothesis must be uniform across any subset of evidence (which it is likely not). Although Bayesian process tracing does not have an explicit method for theory generation via "soaking and poking,"[33] the push to break down the line between inductive and deductive reasoning combined with the claim that new evidence is not uniquely valuable has clear implications for this aspect of research.[34] Say a researcher stumbles upon a new puzzle. At the outset, the number of hypotheses is effectively infinite. Given the temporal, financial, and cognitive costs of reading the full set of relevant evidence $\mathbb{E}$, she will begin by reading some subset of $\mathbb{E}$ to get a sense of what accounts for the outcome. At a certain point, she has to start placing bets on some hypotheses over others based on what she has seen initially. For example, among E_{1-10} perhaps five pieces of evidence inspire H_a, another three inspire H_b, and two inspire H_c.

Thus, the first problem with asserting the irrelevance of timing arises as a function of costly search. The evidence researchers encounter *initially* will affect which hypotheses they devise, refine, and test. Scholars effectively face the multi-armed bandit problem in the context of hypothesis generation.[35] Given limited resources, they are forced to optimize between exploration (reading more material to devise additional hypotheses) and exploitation (settling on a few

---

30 Fairfield and Charman (2019, 160).
31 Fairfield and Charman (2019, 157) and Bennett (2015a, 293).
32 Fairfield and Charman (2019, 156).
33 Bennett (2015a, 276).
34 Furthermore, according to Beach and Pedersen, "empirics-first research is *also* scientific in Bayesian reasoning" (2019, 175).
35 The multi-armed bandit problem represents the trade-offs inherent when learning is costly. The classical example is a person with limited funds facing a row of slot machines with unknown payoffs. Initially, he will play a variety of machines. At a certain point, however, he will begin favoring the machines that had early payoffs, rather than exploring other machines.

PA

hypotheses based on initial material and following the corresponding leads). As such, "what we knew and when"[36] plays a role in research from the outset.

In the current example, the researcher will assign relatively high priors to H_a, slightly lower priors to H_b and even lower ones to H_c. In short, what scholars choose as their "training data" (i.e., the *initial* set of evidence examined—and thus, the order in which scholars observe evidence) will affect the probability of choosing a given hypothesis, the prior probability assigned to it, and the priors assigned to any hypothesis devised later (since ad hoc hypotheses should be penalized).[37] A different sampling of evidence may have altered the priors substantially, and while Bayesians argue that prior probability functions will converge in the long run,[38] Putnam reminds us that for our results to be valid, convergence must "be reasonably rapid,"[39] since "in the long run, we'll all be dead."[40] To embrace a method that relies on convergence across scholars, methodologists must examine when and whether convergence happens, and what goes wrong if it does not.[41]

3.2 Hypothesis Refinement

One of the most touted benefits of the Bayesian approach is that it provides a systematic framework for not just updating confidence in hypotheses, but also refining the hypotheses themselves as researchers "move back and forth between theory development and data."[42] However, valuing iterative refinement appears to directly contradict the global irrelevance of timing. Iteration, by definition, is about revising and updating our beliefs as we move through additional evidence over time. Yet, arguments that degrade "new" evidence—claiming it has no analytically distinct benefits over what researchers have already seen—undermine the value and process of iterative refinement, thereby creating logical inconsistencies in the method.

In the context of hypothesis refinement, disregarding the role of *new* evidence gives rise to what is known mathematically as a "stopping problem." To illustrate, say we start with a hypothesis, H. After analyzing evidence E_1, we slightly modify our hypothesis to H'. Then, we move on to evidence E_2 and modify our hypothesis again to H''. If indeed we can use $E_{1,2}$ to *test* H'', how do we know when to stop analyzing data? While E_1 and E_2 certainly support H'', it seems only sensical to continue testing with *additional* evidence to not only assess its accuracy, but to assess its *stability* as well. New evidence can establish whether we have arrived at a stable hypothesis, or whether further refinements are warranted.[43] If, alternately, the same piece of evidence can inspire, refine, and test, researchers are susceptible to reporting unrefined hypotheses by stopping their analyses too soon.[44]

3.3 Hypothesis Testing

Turning to hypothesis testing, I evaluate how arguments 1 and 2 affect inference. In light of the claim that new evidence has no distinct value, it follows that analysts can test a hypothesis using the evidence that inspired it.[45] On the one hand, this assertion forces us to think more critically about the inferential value of "inspiring" evidence. On the other hand, it downplays and

---

36 Fairfield and Charman (2019, 160).
37 Fairfield and Charman (2019, 162).
38 Bennett argues that "with enough…evidence…differences in researchers' priors should 'wash out,' and posteriors should converge to similar values" (Bennett 2015a, 289).
39 Putnam (1981, 3633).
40 Keynes (1989 [1923], 80).
41 It is also worth noting that the discipline does not have a strong tradition of replication, which raises questions about whether we can expect this debate and convergence to take place in practice.
42 McKeown (1999, 180); Fairfield and Charman (2019, 155); Beach and Pedersen (2019, 174–175).
43 Bennett (2015a, 27–28).
44 Crucially, Bennett diverges from Fairfield and Charman's view and highlights the importance of additional evidence in addressing the stopping problem (2015a, 27). While his approach is not susceptible to the problem highlighted here, scholars new to the qualitative Bayesian literature are met with contradictory advice and no way of adjudicating between the two views.
45 Fairfield and Charman (2019, 160).

misrepresents value of "new" or additional evidence in Bayesian analysis. Ultimately, the attempt to break down the line between inductive and deductive stages of research is taken too far.

Providing a framework for incorporating inspirational evidence into our analyses is a positive contribution. Falling in line with Lieberman's recommendation to move toward disciplinary norms that value (and publish) descriptive and inductive research,[46] the push to recognize the value of "old evidence" calls into question the value of time stamping and pre-registration. As Lieberman argues, "if we take the idea of pre-registration too far...we will surely crowd out the important inductive work upon which scientific discovery depends."[47] For example, the best piece of evidence (in terms of its clarity, specificity, or just interestingness) might be the one that inspired the final version of a hypothesis or theory. While the researcher should likely conduct additional research to ensure the hypothesis's *stability*, pretending that old evidence has no analytic value does little to improve our inferential validity.[48]

However, this contribution is overshadowed by multiple problems. First, proponents of using inspirational evidence to evaluate hypotheses fail to distinguish between *supporting* a hypothesis (which inspirational evidence can do), and *testing* a hypothesis (which it cannot). To test something is, by definition, to expose it to strain—and inspirational evidence will not strain the hypothesis derived from it. Returning to the hypothesis-generation example, the scholar may treat E_{1-10} as sufficient to conclude that H_a outperforms H_b and H_c. While this conclusion may be true, the tactic of using old evidence to support *and* test it is only reliable if the distribution of support for all hypotheses in *any subset* of the evidence is equivalent to the distribution of support across the *full set* of evidence, $\mathbb{E}$. In reality, support for H_a may be systematically clustered in E_{1-10}.[49] In mathematical terms, using the same subset of evidence to inspire and test a hypothesis may lead researchers to choose a hypothesis that represents a local maximum, but report it as though it constitutes a global one.

Another problem is that these recommendations imply a false trade-off between revaluing old evidence and devaluing new evidence. Yet, this is not a zero-sum game. Contrary to the argument underlying this claim,[50] a distinct theory-testing stage is not solely a frequentist notion, and "new" evidence plays a core role in Bayesian inference. Even in the optimal research setting, examining additional evidence beyond the "training set" narrows the band of uncertainty around a researcher's beliefs. In less ideal situations, examining additional evidence can alert the researcher to counterevidence or necessary modifications to their hypotheses. By downplaying the value of new evidence, those recommending this approach undermine the process and contribution of iterative research in the first place.

The second claim pertinent to hypothesis testing is that the results should be insensitive to the order in which the evidence is examined.[51] If this assertion holds in practice, then not only does the Bayesian approach obviate the timing considerations that underly the DA-RT recommendations, but it also constitutes a remarkable tool for inference by ostensibly shielding researchers to temporal biases like serial-position effects.[52]

The outstanding problem with this claim (that the order of evidence does not matter) and the previous one (that new evidence has no unique benefits) is that a significant portion of Bayesianism's added-value hinges on researchers' abilities to overcome known cognitive

---

46 Lieberman (2016).

47 (2016, 1064).

48 Furthermore, scholars using quantitative observational data often have knowledge about one or two cases that inspired the project, yet we do not enjoin them to remove their "pet case" as a row from their dataset.

49 For example, the evidence may have came from similar types of sources, or from a single point in time.

50 Fairfield and Charman (2019, 155).

51 Bennett (2015a, 293); Fairfield and Charman (2017, 365) and (2019, 157).

52 Serial-position effects are cognitive biases that dictate how salient information is deemed as a function of the order in which that information was presented. In its most common forms, subjects overweight or only remember the first or last items in a series (Crano 1977).

limitations. Advocates of the method instruct us to put inspirational evidence out of our minds to derive priors[53] and to ensure that the order in which we examine evidence does not affect our final probabilities.[54] While this advice is well intentioned, researchers should be wary of the method until Bayesian methodologists provide evidence that the recommendations can be reliably implemented. Psychological research suggests that the reality of human cognition is one in which order matters.[55] It is at best naïve—and, at worst, inciting of bias—to argue that the order in which we incorporate evidence should not matter because "the rules of conditional probability mandate [it]."[56] The question is not whether we *should* be able to disregard sequence; the question is whether we *can*. Thus, to the extent that the main contributions of the Bayesian approach require us to overcome these biases, advocates of the method must demonstrate that researchers are capable of arriving at both consistent and accurate conclusions irrespective of the sequence in which they see evidence.[57]

3.4 Reporting Hypotheses and Results

Finally, I evaluate how claims about the irrelevance of timing affect the final product: that is, how researchers structure and report their findings. Breaking from the intuitive appeal of the Bayesian approach, Fairfield and Charman argue that "reporting temporal details about how the research process unfolded" affords "few analytical benefits."[58] Instructing scholars to eschew reporting "what they learned and when" will likely create two problematic ambiguities in published research.

First, this recommendation needlessly reduces transparency and thwarts readers' chances of catching reasoning errors. Proponents of the method argue that the utility of Bayesian process tracing is to formalize iterative research while subjecting the process of iteration and choices of probabilities to scrutiny via peer review.[59] Indeed, the crux of Bayesian transparency is that reviewers and future researchers can challenge and amend poorly-chosen or biased priors. Yet, if researchers succumb to serial-position effects or make other temporally based errors without accounting for what they learned and when they learned it, their readers will lack the information needed to assess whether the chosen probabilities were influenced more by timing than reality.

The second issue is one of practicality and readability. Even if the order in which a researcher analyzes her evidence proves entirely irrelevant to the probabilities she assigns and conclusions she draws, why encourage her to disregard the sequence in her narrative? If we do not structure our write-ups to broadly map onto the sequence of iterative updating, what is the alternative way of structuring them? While it is worth noting that the order in which researchers analyze evidence should not affect their conclusions, enjoining researchers not to report that order only further reduces the transparency of the process and leaves them without an alternative structure that improves on using analytic sequence as a narrative anchor.

4 Claim 3: The Bayesian Approach Fully Engages Rival Explanations

The third claim I address—and the motivation for using the odds-ratio form of Bayes' rule—is that Bayesian process tracing enables scholars to more explicitly engage rival explanations.[60] Fully engaging alternative explanations is crucial for making valid inferences and is a hallmark of process tracing in all its forms.[61] While the recommendation to engage alternative hypotheses

---

53 Fairfield and Charman (2019, 160).
54 Bennett (2014, 293); Fairfield and Charman (2017, 365).
55 Asch (1946); Crano (1977).
56 Fairfield and Charman (2019, 160).
57 On the other side of this coin, Bayesian proponents must also demonstrate that potential reviewers can reliably and accurately catch errors if they do exist.
58 (2019, 156).
59 Bennett (2015a, 289); Fairfield and Charman (2017, 377), (2019, 163).
60 Bennett (2015b, 231); Fairfield and Charman (2017, 373), (2019, 155).
61 George and Bennett (2005); Brady (2006); Hall (2006); Bennett (2010); Collier (2011).

is well-received, the Bayesian method for inference relies on the often incorrect assumption that all rival hypotheses are mutually exclusive: that is, they cannot simultaneously be true.[62] Mutual exclusivity is a core modeling assumption in the Bayesian framework, and the question I ask here is, *what goes wrong if it does not hold?*

4.1 Treatment of Rival Hypotheses under Bayesianism

In current formulations of the method, Bayesian scholars propose two hypotheses, parenthetically note that they assume mutual exclusivity among them, and proceed with the analysis.[63] For example, when Fairfield and Charman apply the framework to Kurtz's state-building research, the authors write: "We wish to ascertain whether the resource-curse hypothesis, or the welfare hypothesis (assumed mutually exclusive), better explains institutional development in Peru."[64] While proponents of the method are consistent in noting this assumption, they do not instruct scholars on how to assess whether mutual exclusivity holds or what goes wrong when multiple hypotheses may work simultaneously (but are analyzed as though they do not). Despite an established typology of relationships among alternative explanations and a corresponding expansion to Bayes' rule to accommodate nonexclusivity,[65] none of the proponents of Bayesian process tracing has incorporated these insights into the method.

4.2 Problems and Implications

The Bayesian approach is not equipped to handle the range of forms hypotheses and evidence tend to take. The problems are both substantive and mathematical, which together result in a method that is more limited than any of its proponents acknowledge. Substantively, the method encourages an oversimplification of the world by sidestepping the frequency with which two causal factors together bring about an outcome. Indeed, Bayesian methodologists frequently use examples that likely violate the mutual exclusivity assumption.[66] By pitting nonexclusive explanations against one another, the Bayesian approach forces hypotheses to take a strong form in which evidence supporting one hypothesis is interpreted to necessarily undermine any other. In reality, many hypotheses implicitly take the form "this matters, too," but the literature is unclear on how to proceed in those instances, beyond acting as though they are exclusive anyway.

Mathematically, when two nonexclusive hypotheses are treated as though exclusivity holds, nearly every term in the equation is inaccurate. To motivate this discussion, imagine comparing two hypotheses that could jointly be true: whether greed ($H_{\$}$) or grievance ($H_{\odot}$) motivates participation in rebellion (given by Equation (2)). Nothing about being upset with the government precludes greediness and rebels may easily exhibit both traits.

$$\underbrace{\frac{P(H_{\$}|EI)}{P(H_{\odot}|EI)}}_{\text{Posterior}} = \underbrace{\frac{P(H_{\$}|I)}{P(H_{\odot}|I)}}_{\text{Prior}} \times \underbrace{\frac{P(E|H_{\$}I)}{P(E|H_{\odot}I)}}_{\text{Likelihood}} . \tag{2}$$

The first problem lies in the formulation of the prior: the probability that greed ($H_{\$}$) explains participation over the probability that grievance ($H_{\odot}$) explains participation. Since greed and grievance can operate together, the prior disregards any extent to which the world is represented

---

62 As Zaks (2017) illustrates, this assumption is pervasive across the process-tracing literature, including, but not limited to the Bayesian variant.

63 Bennett (2014, 47), Fairfield and Charman (2017, 367), Fairfield and Charman (2019, 159) and Barrenechea and Mahoney (2019, 458) use either footnotes or parenthetical statements to note that they assume mutually exclusive hypotheses. While Bennett has a footnote explaining that mutual exclusivity may not hold, the discussion does not engage what goes wrong if the assumption fails.

64 Fairfield and Charman (2019, 159), discussing Kurtz (2009).

65 Zaks (2017, 354).

66 Returning to the example above, it seems plausible that a country may simultaneously exhibit mineral resource dependence and absence of welfare and *both* may contribute to curtailed institutional development.

by the quantity $P(H_\$ \cap H_\odot)$. [67] If researchers proceed as though mutual exclusivity holds, they must first justify the decision theoretically (since "*only greed*" is a very different hypothesis from "greed affects participation"), and compute the numerator and denominator as $P(H_i|I) - P(H_\$ \cap H_\odot)$. This approach raises questions about how the quantity $P(H_\$ \cap H_\odot)$ is assessed and how researchers can identify whether they are incorrect.

The mutual exclusivity assumption also plagues the likelihood ratio (the odds of observing a piece of evidence in world $H_\$$ versus world $H_\odot$). Researchers must do extra legwork to justify that evidence of greed is evidence in favor of *only* greed (rather than greed and grievance together). This task may simply involve finding separate, but disconfirming evidence for the alternative hypothesis—though the method's proponents do not mention this requirement. Without acknowledging and subtracting the possibility of being in the overlap space, estimates for the numerator are likely to be artificially inflated, and estimates for the denominator are likely to be artificially attenuated, thereby introducing confirmation bias.

The final problem lies in the denominator of the likelihood ratio and derives from a related, yet unacknowledged, assumption about the nature of evidence. Specifically, this quantity only returns valid results when every piece of evidence is relevant to all hypotheses. In the context of nonexclusivity, however, evidence that supports one hypothesis is not likely to affect the plausibility of the other. For example, an interview with a former rebel who expressed a desire to profit from illicit diamond mining ($E_\diamond$) does not tell us anything about grievance. If a piece of evidence is only relevant to greed ($H_\$$), how can scholars derive or interpret $P(E_\diamond|H_\odot)$? Mathematically, $E_\diamond$ and $H_\odot$ are independent events; consequently, the denominator reduces to $P(E_\diamond)$ (i.e., the probability of finding this piece of evidence anywhere).[68] Even if we have the ability to assess the overall probability of observing a piece of evidence unconditional on any hypothesis, the likelihood ratio is no longer comparing what it purports to compare.

The implications of this problem extend beyond a misused conditional probability. If the evidence supports one hypothesis but is unrelated to the other, this equation introduces *disconfirmation* bias—in which neutral evidence acts as an undue penalty against the unrelated theory. For example, since $E_\diamond$ supports $H_\$$, then the probability of observing $E_\diamond$ conditional on $H_\$$ is necessarily greater than $P(E_\diamond)$ overall. Consequently, evidence that is *substantively* uninformative about grievances will cause researchers to conclude not only that greed is the *more plausible* explanation (which is fine because we found evidence supporting it), but also that grievance is *less plausible* despite not having found evidence against it. In reality, both hypotheses still may be true.

All analytic techniques have assumptions, limitations, and trade-offs. The responsibility falls to the methodologist to make known the technique's limits and the consequences of proceeding beyond them. These guidelines for appropriate use are absent in the current literature. The infrequency with which mutual exclusivity holds in practice belies qualitative Bayesians' call for wide adoption of the method. Moreover, the breadth of traction qualitative Bayesian analysis has gotten in the discipline in the absence of solutions for dealing with nonexclusive hypotheses puts scholars at risk for adopting a method that is inappropriate for the majority of our questions.

5 Claim 4: The Bayesian Approach Counteracts Biases

The final claim I evaluate is that Bayesian reasoning helps prevent two cognitive biases plaguing qualitative research: confirmation bias and ad hoc hypothesizing.[69] Guarding against

---

67 For a more extensive discussion of overlap in the probability space in the context of deriving Bayesian probabilities, see Zaks (2017, 353–354).

68 Conditioning a probability on an independent event does not alter the probability. Consider the following question: *what's the probability of a fair coin landing on heads given that it's Tuesday?* Since the distribution of outcomes on a fair coin is unrelated to the day of the week, the answer is still 0.5.

69 McKeown (1999, 183); Bennett (2015b, 238); Fairfield and Charman (2019, 155).

confirmation bias and ad hoc hypothesizing is crucial to ensure a theory's validity beyond evidence examined. The question I raise here is whether the Bayesian approach is well-suited to being the shield.

5.1 Reducing Confirmation Bias

According to its proponents, Bayesian process tracing reduces confirmation bias in two ways: using likelihood ratios to ensure researchers attend to competing hypotheses and using conditional probabilities.[70] Likelihood ratios are supposed to reduce confirmation bias by "precluding the pitfall of restricting attention to a single hypothesis."[71] As I demonstrate in the previous section, however, their utility is limited to testing mutually exclusive hypotheses. In short, while good inference may "always involve comparing hypotheses,"[72] it does not always involve *competing* hypotheses. Without the latter, Bayesian analysis will induce, rather than correct, confirmation bias.

The other mechanism aimed at reducing confirmation bias is the prescription to condition probabilities on "all relevant information available without presuming anything beyond what is known or bringing mere opinions or desires into evaluation."[73] Notwithstanding the intuitive framework for thinking about the impact of evidence on our hypothesis, the prescription to condition probabilities only on what we know objectively is not sufficient to ensure proper execution. In reality, this prescription is no different than telling people not to cherry pick evidence or allow fondness for our pet hypothesis to color our evaluation of it. Beyond the argument that using explicit probabilities forces us to justify our choices and allows other scholars to evaluate them in the review process, the approach lacks any mechanism to *enforce* the recommendations.[74]

5.2 Correcting for Ad hoc Hypothesizing

In a research context where evidence can both inspire and test hypotheses, preventing the tendency to "over-tailor an explanation to fit a particular...set of observations" is a critical and valuable safeguard.[75] The logical Bayesian corrective to ad hoc hypothesizing is to "penalize [the prior on] *complex* hypotheses if they do not provide enough additional explanatory power relative to simpler rivals."[76] Initially, this approach seems reasonable: allow a high likelihood for evidence that inspired it, but assign low prior, since we devised the hypothesis post hoc. Upon further scrutiny, however, this recommendation exposes fundamental problems that compromise both its implementation and effectiveness.

The first problem is that the recommendation conflates "ad hoc" and "complex" without justifying the equivalence or providing guidelines to assess the severity of the infraction. Literally meaning "to this," *ad hoc* hypotheses are constructed to fit particular pieces of evidence. They become problematic when they are tailored to idiosyncratic observations, thereby lacking generalizability. While an ad hoc hypothesis may be "arbitrary or overly complex," focusing exclusively on the latter sidesteps the corrective most needed: testing the ad hoc hypothesis with new evidence to assess whether it holds up to further scrutiny. However, since neither arbitrariness nor excessive complexity are desirable traits for a theory, I engage this recommendation on its own terms and evaluate how Bayesian logic addresses these problems.

---

70 Bennett (2014, 2015a); Fairfield and Charman (2017, 2019).
71 Fairfield and Charman (2019, 161).
72 Fairfield and Charman (2019, 158).
73 Fairfield and Charman (2019, 160–161).
74 Moreover, interscholar disagreements about probabilities in the review process may be a function of bias from one, or the other, or both (or all three) parties in the debate.
75 Fairfield and Charman (2019, 161).
76 The quote is from Fairfield and Charman (2019, 161), yet the recommendation to assign lower priors to retrofitted models (i.e., ad hoc hypotheses) originates in McKeown (1999, 183).

While the Bayesian recommendation to penalize the priors on the complex hypothesis has intuitive appeal, the literature lacks concrete guidelines for three corresponding tasks: how to evaluate relative complexity, how to scale penalties accordingly, and how to assess trade-offs. Fairfield and Charman define a complex hypothesis as one that "invokes many more causal factors or elaborate conjunctions of causal factors."[77] But, is a hypothesis with four causal factors so much more complex than a hypothesis with three that it deserves a penalty and how big of one? How do scholars adjudicate between a simple, yet ad hoc hypothesis, and a complex, but more general alternative devised a priori? For this method to be as useful of a practical tool as is it a metaphorical tool, researchers need a concrete guide to answer these questions.

The value placed on simplicity forces us to ask just how simple our explanations have to be to pass muster, and what simplicity buys us in terms of explanatory power. Take, for example, the democratic peace proposition, which Bruce Russet called "one of the strongest nontrivial and nontautological generalizations that can be made about international relations."[78] On the surface, the hypothesis is simple and elegant: democracies do not go to war with one another. But under the hood, the moving parts are countless: what it takes to be a democracy, what aspects of democracy are salient, how normative and institutional mediators matter for preventing conflict, and so on. The truth of the matter is that Occam's Razor is infrequently well-suited to social phenomena—and until someone demonstrates that the simpler explanation tends to be the right one where politics is concerned, penalizing marginal complexity is unsubstantiated.

Returning to the main prescription—to penalize priors on ad hoc hypotheses—the second problem comes to light when one tries to assess whether an ad hoc hypothesis "*provides enough additional explanatory power* relative to a simpler rival."[79] This recommendation lacks pragmatic guidelines for evaluating what constitutes *enough* explanatory power and it highlights an inherent contradiction in the method. Since ad hoc hypotheses are problematic because they are tailored to fit singular observations, then—contrary to the claim that new evidence has no special status over old evidence—"new" evidence is uniquely valuable for evaluating the scope of the hypothesis's explanatory power.[80] Without new evidence, the Bayesian approach is unable to resolve the fundamental problem with ad hoc hypotheses.

The final problem is that the penalties on ad hoc hypotheses are assigned "relative to simpler rivals."[81] Per the previous section, political phenomena are often (if not always) driven by multiple causal factors.[82] As such, by pitting hypotheses against one another by default—especially without justifying the superiority of simpler explanations—the Bayesian approach runs the risk of pushing scholars away from complete and accurate descriptions of the phenomena they are investigating.

6 Discussion

I conclude with a discussion of the implicit 5th Claim of Bayesian process tracing: that the technique's benefits justify the costs of adoption. To be sure, this claim is inherent to all methodological advancements, and it warrants careful examination given that researchers face trade-offs between breadth and mastery in their training. I evaluate the utility and opportunity costs of Bayesian process tracing (or any methodology) on three dimensions: (1) the extent to which the method delivers on its claims, (2) the extent to which it constitutes an improvement over existing techniques, and (3) the extent to which it minimizes opportunity costs to researchers.

---

77 (2019, 161).
78 Russett (1990, 123).
79 Fairfield and Charman (2019, 161), emphasis added.
80 It is worth noting that Bennett's approach to Bayesian analysis corroborates the role of additional evidence (2015a, 28). The remaining problem is that Bayesian methodologists ought to engage more directly with one another to reconcile these contradictions.
81 Fairfield and Charman (2019, 161).
82 Zaks (2017).

First, to justify adoption at the scale to which qualitative Bayesian methodologists aspire, Bayesian process tracing must live up to its central claims. In their current state, however, the techniques are insufficiently developed to achieve most of the stated goals. While the narrative (as opposed to the explicit mathematical) form of Bayesian analysis provides an intuitive framework for updating our beliefs as we move through evidence, the method (particularly the mathematical version) exhibits major shortcomings vis-à-vis its other objectives. Although it appears to provide a rigorous template for iterative research, methodologists implementing the technique exhibit contradictory and counterintuitive practices when it comes to updating priors as they examine additional evidence (as demonstrated in Section 1, above). Similarly, while the claim that "timing is irrelevant" highlights the importance of taking inspirational evidence into account, taking the claim as far as its proponents advocate introduces more problems than it corrects for. Next, per Section 3 above, Bayesian techniques remain incapable of dealing with nonexclusive "rival" hypotheses, despite claiming to offer more critical adjudication of alternatives. Finally, the problems associated with incomplete treatment of timing and rival hypotheses introduce, rather than correct, cognitive biases. Taken together, these shortcomings are likely to compromise the analytic transparency that the method is purported to contribute. Thus, while the goals of Bayesian process tracing are important, much work remains to be done before the method lives up to the mission.

The second requirement to justify adopting a new method is that it should improve upon existing techniques.[83] Qualitative Bayesian methodologists have gone to lengths to *replicate* seminal studies—demonstrating that we can reach the same conclusions within a Bayesian framework.[84] Yet, the value added remains elusive.[85] The explicit Bayesian approach, in particular, requires marked start-up costs in the way of training. While extensive methodological training is not inherently problematic, we should only embrace it with the promise that at the end of the tunnel we will be able to improve—rather than just replicate—what came before.

However, the studies consistently referenced and upheld as the most exemplary use only the most basic form of process tracing—the analytic narrative—and none of those who replicate them demonstrates where the Bayesian approach improves the conclusions. Indeed, two of the three cited do not even refer explicitly to "process tracing" in their books. The fact of the matter is that Schultz and Wood and Tannenwald are excellent data collectors, analyzers, and writers—skills that consistently prove to be the most central assets to good (and transparent) process tracing. Until Bayesian proponents can demonstrate where their method reveals new conclusions or more nuanced inferences, the costs of adoption will continue to outweigh the benefits.

The third and related requirement to justify wide adoption is that Bayesian methodologists must demonstrate not only the analytic purchase of the technique, but moreover, they must justify the opportunity costs of investing in it. As I mention above, the start-up costs are significant: one must learn the basics of process tracing, how to use Bayesian math, and—according to Fairfield and Charman—they should develop an intuition for using logarithmic scales to think about how to assign and scale probabilities.[86] While many analytic techniques demand researchers' time to learn both the tools and intuition behind them, we must ask whether the quality of the research

---

83 Improvement may refer to correcting bias or revealing more nuanced inferences than current techniques. Alternately, it may mean replicating existing studies, but with less effort—thereby introducing efficiency into the research process.

84 Bennett (2015a) and Barrenechea and Mahoney (2019) replicate part of Nina Tannenwald's (2007) analysis. Fairfield and Charman (2017, Appendix A) replicate one of Fairfield's earlier substantive articles.

85 It is worth noting that this issue extends to most of the process-tracing advancements developed over the years. Notwithstanding the myriad developments in the methodology, process-tracing scholars consistently refer to the same handful of exemplary studies (Schultz (2001); Wood (2003); Tannenwald (2007)) to demonstrate how new methods lead to the *same* results. See Zaks (2017) for an example of demonstrating how the new framework introduced leads to more nuanced conclusions in the replication.

86 Fairfield and Charman (2017, 372–373).

that results would be worse, equal, or better than the research that would result if researchers had other training in its place.

In the context of qualitative research, scholars have a lot more access to training in the analysis of data than they do in the research processes that get them the data in the first place. But the process of research and the processes we are researching are inextricable. Researchers would likely yield greater benefits from intensive training in ethnographic, interview, and sampling techniques; understanding the politics and biases associated with archival work; or even just additional and specialized language training needed to conduct research on a specific topic.[87] As I mention above, the qualitative work upheld as exemplary is consistently that which exhibits great skill in gathering evidence, rather than cutting-edge techniques in analyzing it. The vast array of critical skills to which researchers could dedicate their finite time and resources raises additional questions about whether the investment in learning Bayesian process tracing is worth the costs.

6.1 The Future of Bayesian Process Tracing

I conclude with a discussion of future directions of Bayesian process tracing. This article has demonstrated that, in their current state, qualitative Bayesian techniques are insufficient to achieve most of the method's goals. Beyond providing better justification of the opportunity costs of training, four areas demand further research and refinement before the method can be considered viable.

First, qualitative Bayesian methodologists must provide more extensive and concrete guidelines for how to proceed with iterative updating. The primary motivation of adopting an explicit Bayesian frame for process tracing is that it provides a systematic method of updating confidence in hypotheses as researchers move through evidence. In its current state, however, the literature both lacks guidelines on key questions (e.g., how to proceed with analyses when testing more than two hypotheses) and exhibits contradictory practices (e.g., not using updated probabilities when analyzing subsequent pieces of evidence). As a result, researchers looking to put this method into practice lack a comprehensive and logically consistent template for executing the core task of Bayesian analysis.

The second point to reconcile is the contradictory role of timing. The assertion that "old evidence has no special status relative to new evidence" is effectively the chorus of Fairfield and Charman's most recent article.[88] Unfortunately, the authors go so far in dismissing any value of timing that they create contradictions and missed opportunities. Advocates of the method would do well to walk this dismissal back and critically examine the utility of new evidence for assessing the stability of evolving theories and overcoming problems with ad hoc hypotheses. Furthermore, the value of the Bayesian approach would be more convincing to the extent that its proponents could demonstrate that researchers are generally capable of not making timing errors, and—when these errors do occur—that reviewers are generally capable of catching them.

The third area demanding further research is the process of assigning probabilities in explicit Bayesian analyses. If indeed Bayesianism guards against cognitive biases, advocates of the method have the responsibility to test whether and to what extent biases arise in practice when conjuring and evaluating quantities like the prior and likelihood functions. The method's utility would be considerably more convincing if proponents could show (perhaps experimentally) that researchers with similar background information arrive at similar probability assessments and similar evaluations of others' assessments. In other words, Bayesian methodologists should test whether the rules and recommendations they enjoin researchers to follow in theory are realistic in practice.

---

87 For example, standard language-learning approaches rarely provide an adequate basis for the vocabulary of conflict research, regional colloquialisms and slang, or other technical terms that come up in the context of interviews or archival work.

88 Fairfield and Charman (2019).

Fourth and most pressing, Bayesian methodologists must address the method's limited capacity to deal with alternative hypotheses. Merely stating that one "assumes mutual exclusivity" without explicitly evaluating how hypotheses relate to one another downplays the severity of this modeling assumption and exaggerates the scope of research questions to which this approach is applicable.[89] The bias resulting from inappropriately assuming mutual exclusivity compromises not only the validity of inferences that follow from a Bayesian analysis, but also the broader claims about its capacity to increase transparency and reduce cognitive biases in qualitative research. Qualitative Bayesian methodologists must decide whether the method is best limited to testing mutually exclusive hypotheses, in which case they must instruct scholars how to assess whether mutual exclusivity holds in a given case. Alternatively, they must adopt (or derive new) expansions to Bayes' rule to accommodate the wide scope of relationships among rival hypotheses we encounter.[90]

To conclude, the principles motivating qualitative Bayesian process tracing have undeniable value. As a discipline, we should value iterative research, acknowledge the inferential contribution of "old" evidence, fully engage alternative explanations, and guard against confirmation bias and ad hoc hypotheses. Moreover, in the context of ongoing transparency debates, Bayesian scholars have raised important questions about the utility of time-stamping evidence and pre-registering research designs—yet they have not sufficiently established that the Bayesian approach is the best way forward. Ultimately, I argue that while the Bayesian approach is a useful metaphor for conducting transparent, iterative research, the specific techniques proposed as the way forward are more cloudy than its proponents currently acknowledge. As such, before we continue the multiyear legacy of training researchers in Bayesian process tracing, it is time the method takes two steps back before it takes another step forward.

Acknowledgements

Supplementary material

For supplementary material accompanying this paper, please visit
https://doi.org/10.1017/pan.2020.10.

References

Abell, P. 2009. "Comparative Narratives in Sociological Explanation." *Sociological Methods & Research* 38(1):38–70.

Asch, S. E. 1946. "Forming Impressions of Personality." *Journal of Abnormal and Social Psychology* 41:258–290.

Barrenechea, R., and J. Mahoney. 2019. "A Set-Theoretic Approach to Bayesian Process Tracing." *Sociological Methods & Research* 48(3):451–484.

Beach, D., and R. B. Pedersen. 2012. *Process-Tracing Methods: Foundations and Guidelines*. First ed. Ann Arbor: University of Michigan Press.

Beach, D., and R. B. Pedersen. 2019. *Process-Tracing Methods: Foundations and Guidelines*. Second ed. Ann Arbor: University of Michigan Press.

Bennett, A. 2009. "Process Tracing: A Bayesian Perspective." In *The Oxford Handbook of Political Methodology*, edited by J. Box-Stefensmeier, H. Brady, and D. Collier, 702–722. Oxford: Oxford University Press.

---

89 Zaks (2017).

90 As noted previously, Zaks (2017, 354–355) explicates the process of evaluating nonexclusive rivals in a Bayesian framework and derived an expansion to Bayes' rule to accommodate explicit testing.

Bennett, A. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Second ed., edited by H. E. Brady and D. Collier. Lanham, MD: Rowman & Littlefield.

Bennett, A. 2014. "Process Tracing with Bayes: Moving Beyond the Criteria of Necessity and Sufficiency." *Qualitative and Multimethod Research* 12(1):46–51.

Bennett, A. 2015a. "Disciplining Our Conjectures: Systematizing Process Tracing with Bayesian Analysis." In *Process Tracing: From Metaphor to Analytic Tool*, edited by A. Bennett and J. T. Checkel, 276–298. Cambridge: Cambridge University Press.

Bennett, A. 2015b. "Using Process Tracing to Improve Policy Making: The (Negative) Case of the 2003 Intervention in Iraq." *Security Studies* 24(2):228–238.

Bennett, A., and J. T. Checkel, eds. 2015. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Bennett, A., T. Fairfield, and H. Soifer. 2019. "Comparative Methods and Process Tracing." American Political Science Association Organized Section for Qualitative and Multi-Method Research, Qualitative Transparency Deliberations, Working Group Final Reports, Report III.1 (January 2019). Available at SSRN: https://ssrn.com/abstract=3333405 or http://dx.doi.org/10.2139/ssrn.3333405.

Brady, H. E. 2006. "Toward a Pluralistic Vision of Methodology." *Political Analysis* 14(3):353–368.

Brady, H. E., D. Collier, and J. Seawright. 2010. "Sources of Leverage in Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Second ed., edited by H. E. Brady and D. Collier, 161–199. Lanham, MD: Rowman & Littlefield.

Collier, D. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44(4):823–830.

Crano, W. D. 1977. "Primacy Versus Recency in Retention of Information and Opinion Change." *The Journal of Social Psychology* 101:87–96.

Fairfield, T., and A. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3):363–380.

Fairfield, T., and A. Charman. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17(1):154–167.

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

George, A. L., and T. J. McKeown. 1985. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

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

Hall, P. A. 2006. "Systematic Process Analysis: When and How to Use It." *European Management Review* 3(1):24–31.

Humphreys, M., and A. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109(4):653–673.

Keynes, J. M. 1989 [1923]. *A Tract on Monetary Reform*. Cambridge: Cambridge University Press.

Kurtz, M. 2009. "The Social Foundations of Institutional Order: Reconsidering War and the 'Resource Curse' in Third World State Building." *Politics and Society* 37(4):479–520.

Lieberman, E. S. 2016. "Can the Biomedical Research Cycle be a Model for Political Science?" *Perspectives on Politics* 14(4):1054–1066.

Mahoney, J. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41(4):570–597.

McKeown, T. J. 1999. "Case Studies and the Statistical Worldview: Reivew of King, Keohane, and Verba's Designing Social Inquiry." *International Organization* 53(1):161–190.

Moravcsik, A. 2014. "Transparency: The Revolution in Qualitative Research." *PS: Political Science and Politics* 47(1):48–53.

Putnam, H. 1981. *Reason, Truth, and History*. Kindle ed. Cambridge, UK: Cambridge University Press.

Russett, B. 1990. *Controlling the Sword: The Democratic Governance of National Security*. Cambridge, MA: Harvard University Press.

Schultz, K. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

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

Van Evera, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.

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

Yom, S. 2018. "Analytic Transparency, Radical Honesty, and Strategic Initiatives." *PS: Political Science and Politics* 51(2):416–421.

Zaks, S. 2017. "Relationships Among Rivals (RAR): A Framework for Analyzing Contending Hypotheses in Process-Tracing." *Political Analysis* 25(3):344–362.