

LSE

THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

LSE Research Online

[Tasha Fairfield](#) and Andrew Charman

The Bayesian foundations of iterative research in qualitative social science: a dialogue with the data

**Article (Accepted version)
(Refereed)**

Original citation:

Fairfield, Tasha and Charman, Andrew (2018) *The Bayesian foundations of iterative research in qualitative social science: a dialogue with the data*. [Perspectives on Politics](#). ISSN 1537-5927 (In Press)

© 2018 American Political Science Association

This version available at: <http://eprints.lse.ac.uk/89261/>

Available in LSE Research Online: July 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

**The Bayesian Foundations of Iterative Research in Qualitative Social Science:
A Dialogue with the Data**

Forthcoming, *Perspectives on Politics*
(accepted June 15, 2018)

Tasha Fairfield (LSE)¹
Dept. of International Development
London School of Economics
T.A.Fairfield@lse.ac.uk

Andrew Charman
Dept. of Physics
University of California, Berkeley

Abstract: We advance efforts to explicate and improve inference in qualitative research that iterates between theory development, data collection, and data analysis, rather than proceeding linearly from hypothesizing to testing. We draw on the school of Bayesian “probability as extended logic,” where probabilities represent rational degrees of belief in propositions given limited information, to provide a solid foundation for iterative research that has been lacking in the qualitative methods literature. We argue that mechanisms for distinguishing exploratory from confirmatory stages of analysis that have been suggested in the context of APSA’s DA-RT transparency initiative are unnecessary for qualitative research that is guided by logical Bayesianism, because new evidence has no special status relative to old evidence for testing hypotheses within this inferential framework. Bayesian probability not only fits naturally with how we intuitively move back and forth between theory and data, but also provides a framework for rational reasoning that mitigates confirmation bias and ad-hoc hypothesizing—two common problems associated with iterative research. Moreover, logical Bayesianism facilitates scrutiny of findings by the academic community for signs of sloppy or motivated reasoning. We illustrate these points with an application to recent research on state building.

1. Introduction

In the context of the replication crisis, APSA's transparency initiative, and surrounding debates, scholars have sought to revalue, explicate, and improve inference in qualitative research that proceeds in an inherently iterative manner, where prior knowledge informs hypotheses and data gathering strategies, evidence inspires new or refined hypotheses along the way, and there is continual feedback between theory and data. This iterative style of research, which is common in process tracing and comparative historical analysis, diverges from norms that mandate differentiating and sequencing exploratory (theory-building, inductive) and confirmatory (theory-testing, deductive) stages of research. Theory-testing purportedly requires new data that did not contribute to inspiring hypotheses, and any deviations from a specified research design should be reported.¹ Furthermore, theory testing is generally granted higher status.²

Advocates of iterative qualitative research have suggested the key to enhancing its status and improving inference lies in finding ways to conform to norms of differentiating exploration from confirmation and testing theory with new evidence. Scholars have called for greater transparency about analytical sequencing³ and advocate various mechanisms for keeping track of when a hypothesis was devised relative to specific stages of data collection, including pre-registration,⁴ online logging,⁵ or time-stamping data as “used” vs. “unused” throughout the research process.⁶ Meanwhile, a recent APSA joint-committee proposal for a political science registry asserts: “The basic analytical difference between induction and testing is as relevant to qualitative analysis as to quantitative. ...The clearest evaluation of explanatory or theoretical propositions derives from a new set of observations.”⁷

We present a different view of iterative research that is grounded in Bayesian “probability as extended logic” from the physical sciences,⁸ where probabilities represent rational degrees of belief in propositions given the inevitably limited information we possess. From a logical Bayesian perspective, prescriptions for separating exploratory from confirmatory research build on false dichotomies between old vs. new evidence and inductive vs. deductive reasoning. Theory testing—understood in Bayesian terms as inference to best explanation using probabilistic reasoning—takes *all* evidence into account, regardless of whether it was known to the investigator at the time hypotheses were devised; new evidence has no special status relative to old evidence. Scientific inference invariably entails a “dialogue with the data,” where we go back and forth between theory development, data collection, and data analysis, rather than a linear sequence from hypothesizing to testing.

Our perspective highlights and aims to resolve an underlying tension in efforts to understand and improve qualitative research. On the one hand, much of the best such research implicitly and intuitively, albeit not consciously, approximates the logic of Bayesian reasoning. On the other hand, proposals advocating crisp delineations between exploratory and confirmatory research are grounded in the frequentist inferential framework that still underpins most large-N analysis—a framework that is inapplicable to small-N case-study analysis.

¹ Humphreys et al. 2013, 1; Monogan 2015.

² Bowers et al. 2015, 7; Lieberman 2016, 1057; Jacobs 2018.

³ Yom 2015, 11; Büthe and Jacobs 2015, 55

⁴ Bowers et al. 2015; Jacobs 2018

⁵ Yom 2018

⁶ Kapiszewski et al. 2015b

⁷ Bowers et al. 2015, 15

⁸ Cox 1961, Jaynes 2003

Whereas separating theory-building from theory-testing is imperative within frequentism, it is unnecessary for Bayesian inference.

Accordingly, this piece aims to make two central contributions. First, we advance efforts to revalue iterative research by elucidating its Bayesian foundations and thereby providing a solid methodological basis that has been lacking in the qualitative methods literature. Second, we explicate the safeguards Bayesianism provides against confirmation bias and *ad-hoc* hypothesizing, which make firewalls between theory building and theory testing unnecessary. We therefore argue that time-stamping and pre-registration (binding or non-binding) are not useful tools in qualitative research, regardless of the practical (in)feasibility of these approaches in particular research programs (e.g., analysis of existing historical data vs. generation of original data through expert interviews). We hope this piece will help inform discussion among multi-method and qualitative scholars on the nature of inference in case-study research, as well as the relative costs and *analytical* benefits of measures that have been suggested for improving research transparency, beyond advocating transparency for transparency's sake.

We begin by overviewing the trajectory of methodological thinking on iterative research and situating our contribution within recent work on Bayesian process tracing (§2). We then introduce the “logical” approach to Bayesian probability (§3). We clarify how this framework differs from the frequentist paradigm, and we elucidate fundamental tenets of logical Bayesianism that mitigate the need for distinctions between exploratory and confirmatory research. The key lies in recognizing that the terms “prior” and “posterior,” as applied to our degree of belief (or confidence) in whether a proposition is true or false, are not temporal notions. Instead, they are purely logical concepts that refer to whether we have incorporated a given body of evidence into our analysis via Bayes' rule. Section 4 illustrates these points with an application to qualitative research on state building.

Section 5 considers potential concerns regarding our arguments that within logical Bayesianism, there is no need to keep track of what the investigator knew when and that “old” evidence is just as good as “new” evidence for assessing rival hypotheses. Our response emphasizes that Bayesian probability in and of itself provides a framework for rational reasoning in the face of uncertainty that simultaneously helps inoculate against cognitive biases and opens analysis to scrutiny by other scholars for signs of such pitfalls. While there are no magic bullets for ensuring and signaling honest and unbiased assessments of evidence in practice, drawing on Bayesian reasoning more consciously in qualitative research, discussing rival explanations more explicitly, and openly addressing observations that run counter to overall conclusions can help further those goals. While scholars need familiarity with the basics of Bayesian probability to implement these suggestions, sophisticated technical training is not necessary to begin improving intuition and inference.

2. Perspectives on Iterative Research

Iterative research has a long tradition in social science. Classic methodological discussions include Glaser and Strauss's work, which emphasizes jointly collecting and analyzing data while developing and refining theory and concepts. Yet these authors largely describe their goal as theory building—not theory testing, which entails “more rigorous approaches” that “come later in the scientific enterprise.”⁹

⁹ Glaser and Straus 1967, 103

Differentiating between theory building and testing remains prevalent even in literature that questions KKV's application of standards from large-N statistical inference to case studies. Ragin expressly criticizes KKV's assertion that "we should not make it [our theory] more restrictive without collecting new data to test the new version of the theory,"¹⁰ but his response stops short of providing a methodological rationale; Ragin simply notes the practical infeasibility of KKV's prescription when "the number of relevant cases is limited by the historical record to a mere handful."¹¹ Brady and Collier's groundbreaking volume stresses the contribution of inductive research to theory innovation.¹² But in emphasizing tradeoffs between different objectives, the volume leaves the dichotomy between theory building and theory testing largely intact.

Similarly, contemporary process-tracing literature retains language that discriminates between induction and deduction. Authors refer to inductive vs. deductive process tracing,¹³ theory-building vs. theory-testing process tracing,¹⁴ and similar variants.¹⁵ Even when acknowledging that process tracing in practice involves a complex combination of theory construction and evaluation, these modes are still treated as analytically distinct,¹⁶ and ideally sequential, where "inductive discovery is followed by deductive process tracing" using "evidence independent of that which gave rise to the theory," (quoting Bennett and Checkel).¹⁷

The relationship between theory building and theory testing is receiving renewed attention in the context of debates over transparency. Yom seeks to elevate the status of disciplined "inductive iteration" while highlighting "truly destructive" practices like "data mining, selective reporting, and ignoring conflicting results."¹⁸ Yet Yom's emphasis on "transparency in practice," which calls for scholars to report when they "had to reconceptualize a causal mechanism as new information comes to light, ...tighten a theoretical argument..., or rewrite a process-tracing narrative,"¹⁹ essentially falls back on the linear research template he critiques, in that the only rationale for requiring such information about the temporal trajectory of the intellectual process lies in standard prescriptions to test inductively-inspired theory with new evidence—otherwise we are promoting transparency purely for transparency's sake. While we agree that scholars should be forthright when conducting iterative research, we will argue that there are few analytical benefits to reporting temporal details about how the research process unfolded.

In reevaluating the relationship between theory building and theory testing, we take inspiration not only from the physical sciences but also from early work on Bayesian underpinnings of case-study research. McKeown instigated a pioneering agenda by observing that KKV's statistical world-view clashes with a logic of "folk Bayesianism:"

Researchers...are 'interactive processors.' They move back and forth between theory and data, rather than taking a single pass through the data. ...one can hardly make sense of such activity within the confines of a classical theory of

¹⁰ KKV 1994, 22

¹¹ Ragin 1997, 3

¹² Brady and Collier 2010

¹³ Bennett and Checkel 2015, 7-8

¹⁴ Beach and Pedersen 2013

¹⁵ Mahoney 2015; Bowers et al. 2015, 15

¹⁶ Mahoney 2015, 201-02

¹⁷ Bennett and Checkel 2014, 268. Van Evera (1997, 45-6) offers a dissenting view.

¹⁸ Yom 2015, 4

¹⁹ Yom 2015, 11

statistics. A [Bayesian] theory of probability that treats it as a process involving the revision of prior beliefs is much more consistent with actual practice...²⁰

Subsequent scholarship makes important strides towards applying Bayesian reasoning in process tracing.²¹ But implications of McKeown’s observation about “interactive processing” have not yet been explored. Formal treatments of Bayesian process tracing have been cast in a deductive, theory-testing framing that emphasizes prospective anticipations about evidence we might encounter, without elucidating the importance of inferential feedback and the role played by induction in conjunction with retrospective analysis of data actually obtained.

We build on McKeown’s insights by arguing that logical Bayesianism provides a firm methodological foundation for iterative research. In the apt phrase of astrophysicist Stephen Gull, Bayesian analysis involves a “dialogue with the data.”²² We draw new insights through a continual, iterative process of analyzing data differently and/or more deeply, revising and refining theory, revisiting evidence, asking new questions, and deciding what kinds of additional data to collect. Inference is always provisional, in that theories are rarely definitively refuted or confirmed—they are constantly amended in light of new ideas and new data. In these inferential cycles we never “use up” or “throw away” previous information—Bayesianism mandates learning from accumulated knowledge via conditional probabilities that take into account all relevant known information. Confidence in one proposition depends on what else we know and generally changes when we make new observations. There is no need within logical Bayesianism to temporally sequence inductive and deductive stages of reasoning. Bayes’ rule allows us to move back and forth fluidly between reasoning about empirical implications of hypotheses and drawing inferences about possible causes from observed effects, and Bayesian probability allows us to assess the weight of evidence whether collected before or after formulating hypotheses.

3. Bayesian Logic of Iterative Research

We begin by reviewing conceptual distinctions between Bayesianism and frequentism, the dominant approach to quantitative inference which often informs how qualitative research is evaluated, and introducing the logical school of Bayesianism, which provides a prescription for rational reasoning given incomplete information (§3.1). We then overview the mathematical framework of Bayesian inference (§3.2). Section 3.3 resolves false dichotomies of new vs. old evidence and deductive vs. inductive research by focusing on the logical—not temporal—nature of “prior” and “posterior” probabilities. Section 3.4 discusses safeguards built into logical Bayesianism that help curtail confirmation bias and *ad-hoc* hypothesizing—two potential pitfalls often associated with iterative research that underpin conventional demands for segregating theory building from theory testing.

3.1 Bayesian Foundations

Frequentism conceptualizes probability as a limiting proportion in an infinite series of random trials or repeated experiments. For example, the probability that a coin lands “heads” on a given toss is equated with the fraction of times it turns up “heads” in an infinite sequence of throws. In this view, probability reflects a state of nature—e.g., a property of the coin (fair or weighted) and the flipping process (random or rigged). In contrast, Bayesianism understands

²⁰ McKeown 1999, 180

²¹ Bennett 2015; Humphreys and Jacobs 2015; Fairfield and Charman 2017

²² Quoted in Sivia 2006

probability as a degree of belief based on a state of knowledge. The probability an individual assigns to the next toss of a coin represents her strength of confidence about the outcome after taking into account all relevant information she knows. Two observers watching the same coin flip would rationally assign different probabilities to the proposition “*the next toss will produce heads*” if they have different information about the coin and/or tossing procedure. For example, an observer who has had the opportunity to examine the coin in advance and discerns that it is weighted in favor of “heads” would rationally place a higher probability on that outcome than an observer who is not privy to such information.

The Bayesian notion of probability offers multiple advantages—most centrally: it fits better with how people intuitively reason under uncertainty; it can be applied to any proposition, including causal hypotheses, which would be nonsensical from a frequentist perspective; it is well-suited for explaining unique events or working with a small number of cases, without need to sample from a larger population; and inferences can be made from limited amounts of information, using any relevant evidence (e.g., open-ended interviews, historical records), above and beyond data generated from stochastic processes. These features make Bayesianism especially appropriate for qualitative research, which evaluates competing explanations for complex sociopolitical phenomena using evidence that cannot naturally be conceived as random samples (e.g., information from expert informants, legislative records, archival sources). Strictly speaking, “*frequentist inference is inapplicable to the nonstochastic setting,*” quoting Jackman and Western.²³

The school of Bayesianism we advocate as the foundation for scientific inference—logical Bayesianism—seeks to represent the rational degree of belief we should hold in propositions given the information we possess, independently of hopes, subjective opinion, or personal predilections. In ordinary logic, truth-values of all propositions are known with certainty. But in most real-world contexts, we have limited information, and we are always at least somewhat unsure about whether a proposition is true or false. Bayesian probability is an “extension of logic”²⁴ in that it provides a prescription for how to reason when we have incomplete knowledge and are thus uncertain about the truth of propositions. When degrees of belief assume limiting values of zero (impossibility) or one (certainty), Bayesian probability automatically reduces to ordinary logic.

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 different orders) must produce identical probabilities, and individuals possessing the same information must assign the same probabilities. Cox, Jaynes, and subsequent scholars proved mathematically that if we represent degrees of confidence in the truth of propositions with real numbers and impose these consistency requirements, we are led directly to the standard sum and product rules of probability, which in turn give rise to all other operations for manipulating and updating probabilities.²⁵

The consistency requirements of logical Bayesianism are more demanding than requirements imposed in approaches that draw on the “psychological” or “subjective” school of Bayesianism common in the philosophy of science literature and many Bayesian statistics textbooks for social science. In this latter personalistic approach, rationality requires

²³ Jackman and Western 1994, 413

²⁴ Jaynes 2003

²⁵ E.g., Cox 1961; Jaynes 2003; Gregory 2005

probabilities to be *coherent*, which means that utility-maximizing agents must decline “Dutch Book” bets, where loss is certain. Coherence in turn implies that probabilities satisfy the sum and product rules. But as long as probabilities satisfy these rules, they can be based on pure opinion or whim—whatever happens to motivate an individual to hold some particular subjective degree of belief. Accordingly, within psychological Bayesianism, individuals possessing the same information need not assign identical probabilities.

We will show that the consistency requirements are the key to understanding the powerful methodological foundation that logical Bayesianism provides for iterative research. First, however, we review the mechanics of Bayesian inference.

3.2. Bayesian Inference

Intuitively speaking, Bayesian reasoning is simply a process of updating our views about which hypothesis best explains the phenomena or outcomes of interest as we learn additional information. We begin by identifying two or more alternative hypotheses. The literature we have read along with our own previous experiences and observations give us an initial sense, or “prior” view, about how plausible each hypothesis is—e.g., before heading into the field or the archives, do we believe the median-voter theory is a much stronger contender for explaining levels of redistribution in democracies than approaches focusing instead on the power of organized actors including business associations and social movements? Or are we highly dubious that the median-voter hypothesis provides an accurate explanation for the politics of inequality? As our research proceeds, we ask whether the evidence we gather fits better with one hypothesis as opposed to another. When we have finished collecting data, we arrive at a “posterior” view regarding which hypothesis is most plausible. Bayes’ rule provides a mathematical framework for how we should revise our confidence in a given hypothesis, considering both our previous knowledge and the information we discovered during our research. If we remain too uncertain about which hypothesis performs best after analyzing the data in hand, we may continue our research and collect additional evidence.

Stated more formally, Bayesian inference generally proceeds by assigning *prior probabilities* to salient rival hypotheses. These prior probabilities represent our rational degree of belief (or confidence) in the truth of each hypothesis taking into account all relevant initial knowledge, or *background information* (I), that we possess. Symbolically, we represent the prior probability for hypothesis H as $P(H|I)$. This follows the conventional notation whereby a conditional probability $P(A|B)$ represents the rational degree of belief that we should hold in proposition A given proposition B —that is, how likely is A if we take proposition B to be true. We then consider evidence E obtained during the investigation at hand. The evidence includes all observations (beyond our background information) that bear on the plausibility of the hypotheses. Finally, we employ Bayes’ rule to update our degree of confidence in hypothesis H in light of evidence E :

$$P(H|EI) = P(H|I) \times P(E|HI) / P(E|I) . \tag{1}$$

On the left-hand side of Bayes’ rule, $P(H|EI)$ is the *posterior probability* of H given evidence E and background information I . On the right-hand side, $P(H|I)$ is the *prior probability* discussed above, and $P(E|HI)$ is the *likelihood* of the evidence—the probability of observing evidence E if the hypothesis is actually true. In the denominator, $P(E|I)$ acts as a normalization factor. Bayes’ rule is nothing more than a rearrangement of the product rule of probability:

$$P(AB) = P(BA) = P(A|B) \times P(B) = P(B|A) \times P(A) , \tag{2}$$

where we substitute H and E for propositions A and B , while also explicitly conditioning all probabilities on our background information I .²⁶

Because inference always involves comparing hypotheses, it is easier to work with the odds-ratio form of Bayes' rule:

$$\frac{P(H_i|EI)}{P(H_j|EI)} = \frac{P(H_i|I)}{P(H_j|I)} \times \frac{P(E|H_iI)}{P(E|H_jI)} \quad (3)$$

where the factor $P(E|I)$ conveniently cancels out. The factor on the left-hand side represents the *posterior odds* on hypothesis H_i relative to H_j —namely, how much more plausible one hypothesis is vs. a rival hypothesis in light of the evidence learned as well as the background information we initially brought to the problem. The first factor on the right-hand side is the *prior odds*—the plausibility of one hypothesis compared to the other based only on our background information. For posterior odds and prior odds, we can think in terms of how willing we would be to bet in favor of one hypothesis vs. the other. The second factor on the right-hand side is the *likelihood ratio*—how plausible the evidence is under one hypothesis relative to the other, or in other words, how likely the evidence would be if we assume H_i is true, compared to how likely the evidence would be if we instead assume H_j is true. The odds-ratio form of Bayes' rule states that the posterior odds equal the prior odds multiplied by the likelihood ratio. Bayes' rule tells us that evidence fits better with a hypothesis H_i than an alternative H_j to the extent that H_i makes the evidence more plausible. How much we end up favoring one hypothesis over another depends on both our prior views and the extent to which the evidence weighs in favor of one hypothesis over another.

Assessing likelihood ratios $P(E|H_iI)/P(E|H_jI)$ is therefore the critical inferential step that tells us whether evidence E should make us more or less confident than we were initially in one hypothesis relative to a rival. The likelihood ratio can be thought of as the probability of observing E in a hypothetical world where H_i is true, relative to the probability of observing E in an alternative world where H_j is true. When evaluating likelihoods of the form $P(E|H_iI)$, we must in effect (a) suppress our awareness that E is a known fact, and (b) suppose that H_i is correct, even though the actual status of the hypothesis is uncertain. Recall that in the notation of conditional probability, everything that appears to the right of the vertical bar is either known, or assumed as a matter of conjecture when reasoning about the probability of the proposition to the left of the bar. In qualitative research, we need to “mentally inhabit the world” of each hypothesis (Hunter 1984) and ask how surprising (low probability) or expected (high probability) the evidence E would be in each respective world. If E seems less surprising in the “ H_i world” relative to the “ H_j world,” then that evidence increases our odds on H_i vs. H_j .²⁷ Again, we gain confidence in a given hypothesis to the extent that it makes the evidence we observe more plausible compared to rivals.

Assessing likelihoods entails thinking about how consistent the evidence is with the world of the hypothesis in question. Would the events, decisions, and statements that E represents follow plausibly and naturally in the world of a particular hypothesis? Or would they seem unusual and unexpected, perhaps requiring additional flukes, coincidences, or complex chains of

²⁶ Accordingly, from the right-hand side of equation (2) we have: $P(H|EI) \times P(E|I) = P(E|HI) \times P(H|I)$. Dividing both sides by $P(E|I)$ yields equation (1).

²⁷ Note that the plausibility of the evidence under a particular hypothesis can be extremely small if events could reasonably have unfolded in a multitude of different ways. However, because we are working with likelihood ratios, we need not worry about exactly how (im)plausible the evidence is under H_i or H_j . Instead, we focus on how much more or less likely what did happen becomes under H_i relative to H_j .

intervening events to occur? Here it may be helpful to consider Coleridge’s notion of “willing suspension of disbelief.” When reading a story involving supernatural forces or time travel, for example, the conjured world might strike us as highly implausible—meaning that $P(H|I)$ is very low. But if we nevertheless provisionally accept this hypothesized world, the story should feel compelling and self-consistent, obeying the internal logic of the imagined world—meaning that $P(E|HI)$ for the events that unfold should not be too low compared to other scenarios we can envision.

Elsewhere, we elaborate guidelines for explicit Bayesian analysis in qualitative research, which entails quantifying all probabilities.²⁸ To illustrate how Bayesian logic can be applied heuristically—i.e., without specifying numerical values for probabilities—consider an example drawing on Kurtz’s state-building research.²⁹ We wish to ascertain whether the resource-curse hypothesis, or the warfare hypothesis (assumed mutually exclusive), better explains institutional development in Peru:

H_R = Mineral resource dependence is the central factor hindering institutional development. Mineral wealth makes collecting taxes irrelevant and creates incentives for subsidies and patronage, instead of building administrative capacity.

H_W = Absence of warfare is the central factor hindering institutional development. Without external threats, leaders lack incentives to collect taxes and build administrative capacity for military defense.

For simplicity, suppose we have no relevant background knowledge about state-building in Peru. Since both hypotheses find substantial support in literature on other countries, we might reasonably assign even prior odds. We now learn the following:

E_I = Peru faced persistent military threats following independence, its economy was long dominated by mineral exports, and it never developed an effective state.

Intuitively, E_I strongly favors the resource-curse hypothesis. Applying Bayesian reasoning, we must evaluate the likelihood ratio $P(E_I|H_R)/P(E_I|H_WI)$. Imagining a world where H_R is the correct hypothesis, mineral dependence in conjunction with weak state capacity is exactly what we would expect, and external threats are not surprising given that a weak state with mineral resources could be an easy and attractive target for invasion. In the alternative world of H_W , E_I would be quite surprising; something very unusual, and hence improbable, must have happened for Peru to end up with a weak state if the warfare hypothesis is nevertheless correct, because weak state capacity despite military threats contradicts the expectations of the theory. Because E_I is much more probable under H_R relative to H_W —that is, $P(E_I|H_RI)$ is much greater than $P(E_I|H_WI)$ —the likelihood ratio is large, and it significantly boosts our confidence in the resource-curse hypothesis.

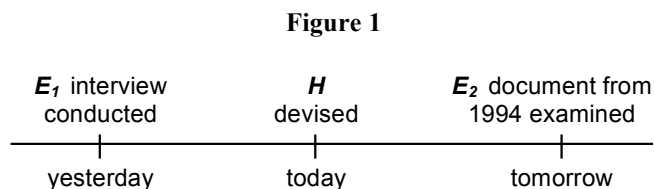
3.3. Prior vs. Posterior Probabilities and Old vs. New Evidence

While testing hypotheses with new evidence is pervasively espoused, distinctions between old and new evidence, and hence exploratory and confirmatory research, are far less consequential within logical Bayesianism compared to frequentism or psychological/subjective Bayesianism. To be clear, “new evidence” refers to information unknown to the scholar before devising the hypothesis—regardless of the historical timing of when that information was

²⁸ Fairfield and Charman 2017

²⁹ Kurtz 2009

generated. For example, in Figure 1, E_1 is old evidence relative to H , whereas E_2 is new evidence, even though E_2 existed before E_1 .



The key to unraveling the false dichotomies lies in understanding that *prior* and *posterior* are not temporal notions—they are *logical* notions. In astro-statistician Tom Loredo’s words:

There is nothing about the passage of time built into probability theory. ... ‘prior probability,’ and ‘posterior probability’ do not refer to times before or after data is available. They refer to logical connections, not temporal ones. Thus, to be precise, a prior probability is the probability assigned before consideration of the data.³⁰

In other words, *prior* and *posterior* refer to degrees of belief before and after a piece of evidence is incorporated into our analysis—not to the timing of when we happened to learn or obtain that evidence. *Prior/posterior* describe idealized states of knowledge without/with specific pieces of evidence included. Hypotheses can contain temporal structuring, and evidence can contain temporal information. However, probabilities themselves carry no intrinsic time-stamps.

These points merit expounding. Recall that within logical Bayesianism, only the data at hand and the background knowledge are relevant for assessing the degree of belief that a hypothesis merits. Nothing else about our state of mind should influence our probabilities. The relative timing of when we stated the hypothesis, worked out its implications, and gathered data falls into this later category of logical irrelevance.

To further stress the logical irrelevance of keeping track of what we knew when, the rules of conditional probability mandate that we can incorporate evidence into our analysis in any order without affecting the posterior probabilities. Using the product rule (2) and commutativity, the joint likelihood of two pieces of evidence can be written in any of the following equivalent ways:

$$P(E_1E_2|HI) = P(E_2E_1|HI) = P(E_1|E_2HI) \times P(E_2|HI) = P(E_2|E_1HI) \times P(E_1|HI). \quad (4)$$

Evidence learned at time one (E_1) may thus be treated as logically posterior to evidence learned at time two (E_2). If in practice conclusions differ depending on the order in which evidence was incorporated, there is an error in our reasoning that should be corrected. Otherwise we have violated the fundamental notion of rationality that lies at the heart of logical Bayesianism (§3.1)—information incorporated in equivalent ways should lead to the same conclusions.

Once we recognize that timing is irrelevant in probability theory, it follows that each step below is logically distinct:

- drawing on evidence E to inspire hypotheses;³¹

³⁰ Loredo 1990, 87

³¹ Inventing hypotheses is a creative process that falls outside probability theory, or any other inferential framework. Quoting MacKay (2006, 346): “Bayes does not tell you how to invent models.”

- assigning prior probabilities to those hypotheses given background information I that does not include E ;
- assessing the likelihood of E under alternative hypotheses to derive posterior probabilities.

Information is neither “exhausted” nor “double-counted” in this process (Appendix A expounds). All relevant knowledge can be sorted as convenient into background information on which all probabilities are conditioned, and evidence that we use to update probabilities.

Psychological/subjective approaches to Bayesianism often diverge from logical Bayesianism on these points, because the former focus on individuals’ personal degrees of belief and how their psychological states evolve over time. Jeffrey’s “probability kinematics” is a prominent example;³² his approach introduces non-standard rules for updating that violate the laws of probability and imply that the order in which evidence is analyzed can matter.³³

In sum, probability theory requires keeping track of *what* information has been incorporated into our analysis, not *when* that information was acquired.³⁴ Time-stamps indicating when hypotheses were composed or when evidence was observed or incorporated are not relevant to scientific inference.³⁵

3.4. Curtailing Confirmation Bias and Ad-Hoc Theorizing

Careful application of Bayesian logic helps guard against confirmation bias and *ad-hoc* hypothesizing in iterative research. We consider these dual pitfalls in turn.

Two common variants of confirmation bias entail overfocusing on data that fit a particular hypothesis and/or overlooking data that undermine it, and focusing on a single favored hypothesis while forgetting to consider whether data consistent with that hypothesis might be as or more supportive of a rival hypothesis. A common recommendation for precluding such biases entails identifying observable implications of rivals as well as the main working hypothesis before gathering data.³⁶ However, this advice can be problematic for two reasons.

First, deducing observable implications beforehand may be infeasible, because any hypothesis may be compatible with a huge number of possible evidentiary findings—just with varying probabilities of occurrence. For qualitative research on complex socio-political phenomena, there is essentially no limit to the different kinds of evidence we might encounter, and there is no way to exhaustively catalogue these infinite possibilities in advance.

Second, anticipating observable implications may foster even greater bias. If we have already elaborated hypotheses to be considered and evidence expected under each, we are now *better situated* to seek out the sorts of evidence that will support our pet theory, compared to a situation where we collect evidence without necessarily anticipating what will support which hypothesis. This caveat embodies classic advice from Doyle’s Sherlock Holmes: “It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts,” (*A Scandal in Bohemia*).

³² Jeffrey 1983

³³ Another salient example is the so-called “new problem of old evidence” (Appendix B).

³⁴ Jefferys 2007

³⁵ In physics, hypotheses often derive support from evidence known long before they were developed; e.g., quantum mechanics was devised to explain known facts about blackbody radiation, atomic stability, and the photoelectric effect.

³⁶ E.g., Bennett and Checkel 2015, 18.

Risks of confirmation bias can be better controlled by conscientiously endeavoring to follow Bayesian reasoning. Tendencies to seek evidence that supports a favored hypothesis, interpret evidence as overly favorable to that hypothesis, and underweight evidence that runs against that hypothesis are counteracted by following Bayesian prescriptions to condition probabilities on all relevant information available, without presuming anything beyond what is in fact known, or bringing mere opinions or desires into the evaluation. Furthermore, remembering that the critical inferential step in Bayesian inference entails assessing likelihood *ratios*, $P(E|H_j I) / P(E|H_k I)$, precludes the pitfall of restricting attention to a single hypothesis—we must always ask whether a given explanation makes the evidence more or less likely *compared to rivals*.

In contrast to confirmation bias, the dual problem of *ad-hoc* hypothesizing involves over-tailoring an explanation to fit a particular, contingent set of observations. This danger underpins calls for distinguishing exploration from confirmation and testing hypotheses with new data. Within logical Bayesianism, however, an *ad-hoc* hypothesis that is too closely tailored to fit arbitrary details of the data incurs a low prior probability, which protects us from favoring it over a simpler hypothesis that adequately explains the data. If an explanation is *ad-hoc*, careful consideration should reveal that it is just one member of a large family of more or less equally *ad-hoc* hypotheses, characterized by multiple parameters or arbitrary choices that must be fine-tuned to the data. Each of these related hypotheses might explain a different set of contingent facts, yet none of them would seem any more credible than the others in the absence of the particular body of observations obtained. Even if the overall prior probability of the family of hypotheses $\{H_1 \text{ or } H_2 \text{ or } \dots H_N\}$ is appreciable, this prior probability must be spread over all of the constituent possibilities, such that the prior for any particular H_i will be small.

Consider an example adapted from astrophysicist Bill Jefferys.³⁷ A stranger at a party shuffles a deck of cards, and you draw the six of spades. We might reasonably hypothesize that this card was arbitrarily selected from a randomly-shuffled deck (H_R). A rival hypothesis proposes that the stranger is a professional magician relying on a trick deck that forces you to draw the six of spades ($H_{6\spadesuit}$). While the likelihood of selecting this particular card is 1/52 under H_R , it is far larger under $H_{6\spadesuit}$. However, $H_{6\spadesuit}$ must be penalized by a factor of 1/52 relative to H_R , because without observing your draw, there would be no reason to predict the six of spades as the magician's forced card. $H_{6\spadesuit}$ should be treated as one of 52 equally plausible related hypotheses whereby the magician forces some other card. Accordingly, our single draw provides insufficient evidence to boost the credibility of $H_{6\spadesuit}$ above H_R .

Logical Bayesianism thus penalizes complex hypotheses if they do not provide enough additional explanatory power relative to simpler rivals, in line with Occam's razor and Einstein's dictum that things should be as simple as possible, but no simpler. In quantitative analysis, this task is accomplished via *Occam factors* that are automatically built into Bayesian probability.³⁸ Appendix C discusses Occam factors in more detail and illustrates how the penalty of 1/52 in our card-draw example emerges when we formally apply Bayesian analysis.

In qualitative social science, the role of Occam factors in penalizing overly-complex explanations was first highlighted by Western.³⁹ Although many authors identify complexity of theories with the number of assumptions or causal factors, there are no universal prescriptions for assessing how *ad-hoc* a hypothesis is in qualitative research. Our recommended stratagem entails scrutinizing new hypotheses to evaluate how much additional complexity they introduce

³⁷ Jefferys 2003

³⁸ Jaynes 2003, 601-07; MacKay 2003, 343-356; Gregory 2005, 45-50

³⁹ Western 2001

compared to rivals. If the hypothesis invokes many more causal factors or very specific or elaborate conjunctions of causal factors, good practice would entail penalizing its prior relative to the rivals, although it can be difficult to judge precisely the tradeoff between simplicity and explanatory power—a point Western also stresses.⁴⁰

In sum, within logical Bayesianism, likelihood ratios help guard against confirmation bias, while priors help guard against *ad-hoc* hypothesizing. These safeguards are absent within frequentism, where hypothesis testing usually focuses on the probability of the data only under the null hypothesis, rather than relative likelihoods under rival hypotheses, and where the concept of probability applies only to data obtained through a stochastic sampling procedure, not to hypotheses. Frequentist inference therefore requires pre-specifying sampling and analysis procedures in detail to avoid confirmation bias, and strictly separating data used in theory-building from data used for theory-testing to prevent *ad-hoc* hypothesizing, whereas such strictures are unnecessary for Bayesian inference.

4. Iteration in Practice

We have argued that within logical Bayesianism, there is no need for firewalls between theory-building and theory-testing, and no need to rely exclusively on “new evidence” when testing hypotheses. All we must do is carefully assign prior odds in light of our background information, and carefully assess likelihood ratios for all relevant evidence under our rival hypotheses. This section illustrates how these points apply to qualitative research by extending the state-building example introduced in §3.2. We make no claims about how Kurtz’s research process unfolded. Instead, we draw on hypotheses and evidence from his published work to show how an iterative dialog with the data can give rise to inferences that are as valid as in a purely deductive approach, where all hypotheses were devised prior to data collection.

After comparing the resource-curse and warfare hypotheses in light of E_1 (military threats, mineral wealth, and weak state), suppose we learn the following:

*E_2 = Throughout the 1880s, Peruvian agriculture relied on an enormous semiservile labor force. When Chile invaded, Peruvian elites were far more concerned that peasants remain under control than they were with contributing to national defense. The mayor of Lima openly hoped for a prompt Chilean occupation for fear that subalterns might rebel. The agrarian upper class not only refused to support General Cáceres’ efforts to fight back, but actively collaborated with Chilean occupiers because of Cáceres’ reliance on armed peasant guerillas.*⁴¹

This evidence might inspire a new hypothesis:

*H_{LRA} = Labor-repressive agriculture is the central factor hindering institutional development. Elites resist taxation and centralized control over coercive institutions, because they fear greater vulnerability to local rebellions.*⁴²

To assess which hypothesis better explains the evidence acquired thus far, we must return to our background information and reassign priors across the new hypothesis set: H_R , H_W , and the inductively-inspired H_{LRA} . We then assess likelihood ratios for the aggregate evidence E_1E_2 .

For priors, strictly speaking we should assess the plausibility of each hypothesis taking into account all information accumulated in previous state-building literature. However,

⁴⁰ Western 2001, 375

⁴¹ Kurtz 2009, 496

⁴² Kurtz 2009, 485

systematically incorporating all of our background information is infeasible in social science. Given practical limitations, one reasonable approach keeps equal odds on H_R vs. H_W but gives H_{LRA} a moderate penalty relative to each rival, thereby acknowledging the novelty of this hypothesis with respect to existing state-building research and anticipating skepticism among readers. Another reasonable option places equal odds on all three hypotheses, considering that H_{LRA} is grounded in a longstanding research tradition originated by Barrington Moore.⁴³ While H_{LRA} is not discussed in state-building literature, labor-repressive agriculture has been identified as a crucial factor affecting other macro-political outcomes including regime type, so *a-priori* we might expect this factor to be salient for state-building as well. Furthermore, although H_{LRA} was introduced *post-hoc* (in light of E_2), it is no more or less *ad-hoc* compared to the rivals—upon inspection, none of the three hypotheses seems appreciably more complex than the others. Each identifies a single structural cause that operates by shaping actors’ incentives.⁴⁴

Turning to the evidence, the easiest way to proceed entails assessing likelihood ratios for H_{LRA} vs. H_R and H_{LRA} vs. H_W . Since the overall likelihood ratio factorizes:

$$\frac{P(E_1 E_2 | H_i I)}{P(E_1 E_2 | H_j I)} = \frac{P(E_1 | H_i I)}{P(E_1 | H_j I)} \times \frac{P(E_2 | E_1 H_i I)}{P(E_2 | E_1 H_j I)} \quad (5)$$

we first consider E_1 and then E_2 .

E_1 moderately favors H_R over H_{LRA} . As explained in §3.2, E_1 fits quite well with the resource-curse hypothesis. However, E_1 is not surprising under H_{LRA} ; a weak state with mineral resources would still be an easy and attractive target for invasion if labor-repressive agriculture were the true cause of state weakness. Nevertheless, resource dependence in conjunction with state weakness makes E_1 more expected under H_R . In contrast, E_1 strongly favors H_{LRA} over H_W : whereas this evidence is unsurprising under H_{LRA} , it is highly unlikely under H_W (§3.2).

E_2 very strongly favors H_{LRA} over each alternative. Neither H_W nor H_R speaks to the nature of agricultural relations, whereas in the world of H_{LRA} , semi-servile labor is highly expected given that Peru has a weak state (E_1).⁴⁵ Furthermore, under either H_W or H_R , the behavior of Peruvian elites described in E_2 would be extremely surprising—we would instead expect them to resist the Chilean incursion (however ineffectively, given state weakness) in an effort to retain control over their territory and mineral resources. In contrast, their behavior fits quite well with H_{LRA} in showing that elites’ concern over maintaining subjugation of the labor force undermined the most basic function of the state—national defense. Of course, we know E_2 fits well with H_{LRA} since the former inspired the latter; however, the critical inferential point is that E_2 is *much more* plausible under H_{LRA} relative to the alternatives. Accordingly, this evidence very strongly increases the odds in favor of H_{LRA} .

Overall, the likelihood ratio (5) strongly favors H_{LRA} over both alternatives. E_2 overwhelms the moderate support that E_1 provides for H_R . And all of the evidence weighs strongly against H_W . Accordingly, H_{LRA} emerges as the best explanation given the evidence acquired thus far. If we begin with a moderate penalty on H_{LRA} , the posterior still favors that hypothesis, although the higher the prior penalty, the more decisive the overall evidence needed to boost the plausibility of H_{LRA} above its competitors.

⁴³ Kurtz 2009, 485

⁴⁴ An argument could potentially be made that H_W is somewhat simpler than the rivals, in that it can be articulated a bit more concisely and invokes a single, direct causal process.

⁴⁵ Here we are conditioning the likelihood of E_2 on E_1 , as equation (5) requires. See Fairfield and Charman (2017, 370-71).

In essence, we have now “tested” an inductively-inspired hypothesis with “old evidence.” What matters is not when H_{LRA} came to mind or which evidence was known before vs. after that moment of inspiration, but simply which hypothesis is most plausible given our background information and all the evidence. Imagine that a colleague is familiar with all three hypotheses from the outset and shares essentially the same background knowledge, but has not seen E_1E_2 . S/he would follow a logically identical inferential process in evaluating which hypothesis best explains the Peruvian case: assessing the likelihood of E_1E_2 under these rival hypotheses. It would be irrational for two researchers with the same knowledge to reach different conclusions merely because of when they learned the evidence.

To further emphasize the irrelevance of relative timing, we do not know from reading Kurtz’s article whether he invented H_{LRA} before or after finding E_2 , but that chronological information would not make E_2 any more or less cogent. Our goal is not to reproduce the order in which the neurons fired inside the author’s brain; it is to independently assess which hypothesis is most plausible in light of the evidence and arguments presented.

Of course “new evidence” is often valuable for improving inferences by providing additional weight of evidence. However, the goal of obtaining new evidence is not to supplant existing evidence that inspired the hypothesis, but rather to supplement that evidence and hopefully strengthen our inference. Information is never intentionally disregarded in logical Bayesianism; any subsequent stage of research following the inspiration of a hypothesis must take all previously-obtained evidence into account through the prior probability on that hypothesis. In our example, E_2 , which inspired H_{LRA} , contributes to the strong posterior odds in favor of H_{LRA} , which would in turn become the “prior odds” when analyzing additional evidence.

5. Anticipated Concerns

Logical Bayesianism is an aspirational ideal that usually cannot be fully realized in practice without approximations. In qualitative social science, some degree of subjectivity inevitably enters when assigning probabilities. There is no mechanical procedure for objectively translating complex, narrative-based, qualitative information into precise probabilities. Despite conscientious efforts to follow Bayesian reasoning, we may still commit analytical errors.

Accordingly, this section considers potential concerns with our argument that qualitative research need not demarcate theory-building vs. theory-testing. Our overarching response draws on the premise that research is not only a dialogue with the data, but also a dialogue with a community of scholars. Knowing the temporal trajectory of authors’ thought processes should not matter to how readers scrutinize inferences. If scholars disagree with an author’s conclusions, logical Bayesianism provides a clear framework for pinpointing the loci of contention, which may lie in different priors and background information, and/or different interpretations of evidence. Bayesian itself, whether applied explicitly or heuristically, lays analysis open for all to scrutinize on its own terms. In contrast, it would be misguided to assume that if authors time-stamp hypotheses and evidence, their analysis is sound, whereas if such information is not reported, inferences lack credibility. Regardless of whether temporal details about the research process are provided, scholars must evaluate hypotheses and evidence with their own independent brainpower.

Our discussion includes guidelines for facilitating scholarly dialogue and improving inferences within a Bayesian framework, while highlighting shortcomings of prescriptions for labeling and/or separating exploratory/inductive vs. confirmatory/deductive research. We address anticipated concerns regarding biased priors, biased likelihoods, and scholarly integrity.

5.1. Biased priors

Concerns: (a) Given psychological difficulties in “getting something out of our mind,” we may be unable to assign priors that are not influenced by what we already know about our data.

(b) Given vulnerabilities to cognitive biases, we may over-fit inductively-devised hypotheses to the evidence without adequately penalizing their priors.

Pre-specifying priors is not a sensible solution. We cannot assess a prior before devising the hypothesis, and once we formulate the hypothesis, all relevant information—both background knowledge and evidence E_{pre} acquired thus far—must inform $P(H|E_{pre}I)$, which serves as the “prior” moving forward. Moreover, whether we evaluate $P(H|I)$ and then the likelihood for the total evidence $E_T = E_{pre} E_{post}$ ultimately collected, $P(E_T|I)$, or whether we update along the way, evaluating $P(H|E_{pre}I)$ and then $P(E_{post}|E_{pre}HI) \times P(H|E_{pre}I)$, the final inference should be the same—consistency checks can help ensure equivalence. The timing of when we assess or record priors is irrelevant.

To guard against subconsciously-biased priors (concern (a)), best practices should include the following. First, describe the most salient background information and explain why it motivates a particular choice of priors. If priors are obviously biased in favor of an inductively-derived hypothesis, beyond what is justified by the background information discussed, readers should notice the discrepancy. For instance, in our state-building example, readers might balk if our prior odds strongly favored H_{LRA} over the well-established resource-curse and warfare hypotheses. Likewise, if a well-known study or salient literature is overlooked, readers will request reconsideration of priors in light of that further background information.

Second, consider conducting the analysis with equal prior odds, which avoids biasing initial assessments in favor of any hypothesis. This approach shifts focus to likelihood ratios, with the aspiration that even if scholars disagree about priors—which will be almost inevitable given that everyone has different background information—we may still concur on the direction in which our odds on the hypotheses should shift in light of the evidence. Third, consider using several different priors to assess how sensitive conclusions are to these initial choices, along the lines of our analysis in §4.

For qualitative research that follows Bayesian logic heuristically, the first guideline entails carefully discussing the strengths and weaknesses of rival explanations based on existing literature, which is common practice. The second guideline entails recognizing that readers may initially view a hypothesis with much more skepticism than the author, such that all parties in the scholarly dialogue should pay close attention to scrutinizing the evidence and the inferential weight it provides. Authors should be conservative with their inferential claims until the weight of evidence becomes strong.

Regarding concern (b), scholarly dialogue again serves as a corrective to sloppy analysis. If an inductive hypothesis manifesting multiple fine-tuned variables or inordinate complexity is granted too much initial credence, readers should notice and demand additional evidence to overcome an unacknowledged or underestimated Occam penalty (§3.4). Beyond the simple advice to treat inductively-devised hypotheses with healthy skepticism, three suggestions can help curtail *ad-hoc* hypothesizing: start with reasonably simple theories and add complexity incrementally as needed; critically assess whether all casual factors in the theory actually improve explanatory leverage; and ask whether the explanation might apply more broadly.

In contrast, reporting the temporal sequencing of the research process in and of itself does not help ascertain how severe an Occam penalty a hypothesis should suffer. The critical point is that a hypothesis that is *post-hoc*—devised after the evidence—is not necessarily *ad-hoc*—arbitrary or overly complex. These are distinct concepts. As argued in §4, H_{LRA} is *post-hoc* (relative to E_2), but not *ad-hoc*, because it is no more arbitrary or complex than its rivals.

5.2. Biased likelihoods

Concern: We may succumb to confirmation bias in overstating how strongly evidence favors an inductively-derived hypothesis.

Suggestions for pre-registration and time-stamping in qualitative research⁴⁶ aim to address these concerns, on the premise that differentiating exploratory from confirmatory analysis allows us to more credibly evaluate inductively-inspired hypotheses. Importing this prescription into a Bayesian framework would entail assigning likelihoods to clues we might encounter before gathering data.

Even in light of human cognitive limitations, we find this approach unhelpful. Although a scholar's prospective assessment of likelihoods for "new evidence" might be less prone to confirmation bias than retrospective analysis of "old evidence," confirmation bias could just as easily intrude when gathering additional evidence—by subconsciously looking harder for clues that favor the working hypothesis and/or overlooking those that do not (§3.4).

Moreover, we reiterate the impossibility of foreseeing all potential evidentiary observations in the complex world of social science. Anticipating course-grained categories of observations is not adequate for specifying likelihoods for any actual, concrete evidence that might fit within that class, because specific details of evidence obtained can matter immensely to likelihoods under different hypotheses. Consider the example Bowers et al. present in their discussion of pre-analysis plans for qualitative research: a government has cut taxes, and we wish to assess hypothesis $H_K = \text{Tax cuts were motivated by an interest in Keynesian demand management}$.⁴⁷ The authors delineate evidence $E = \text{Records of deliberations among cabinet officials about the tax cut show "prominent mention of ... Keynesian stimulus,"}$ and they judge the probability of finding such evidence if H_K is true to be very high. However, E as stated above is too vague to assign a meaningful likelihood in advance. Here are two different clues we might encounter in the records:

$E' = \text{The Finance Minister invokes Keynesian stimulus when explaining the tax cuts to other cabinet members.}$

$E'' = \text{One of the cabinet members comments that tax cuts are consistent with Keynesian stimulus, whereafter discussion is interrupted by derisive jokes about Keynesian economics.}$

Suppose further that the time and attention devoted to these mentions of Keynesiansim are similar for E' and E'' , such that both qualify as instances of E as articulated above, even though they carry very different import. Whereas the likelihood of E' might well be high if H_K is true, the likelihood of E'' certainly is not— E'' would be extremely surprising in a world where H_K is correct.

⁴⁶ Bowers et al. 2015; Kapiszewski et. al. 2015b; Jacobs 2018; Yom 2018, 420

⁴⁷ Bowers et al. 2015, 16-17

Bowers et al. recognize this “problem of precision,” noting that E as defined above “still leaves some things open. Just how prominent do mentions of Keynesian logic have to be...? How many actors have to mention it? What forms of words will count as the use of Keynesian logic?”⁴⁸ However, they underestimate the problem. The issue is not just how many mentions or how many actors or what terms we associate with Keynesianism, but an endless array of other possibilities and nuances that depend on the context and manner in which Keynesianism is discussed. However much additional detail we specify before gathering data, we can always invent—and the real world may well produce—another twist or tweak that matters nontrivially. Despite efforts to anticipate what might surprise us ahead of time, science advances most when evidence surprises us in unforeseen ways.

Jaynes, an outspoken advocate of logical Bayesianism in the physical sciences, reinforces these key points:

The orthodox line of thought [holds] that before seeing the data one will plan in advance for every possible contingency and list the decision to be made after getting every conceivable data set. The problem...is that the number of such data sets is usually astronomical; no worker has the computing facilities needed... We take exactly the opposite view: it is only by delaying a decision until we know the actual data that it is possible to deal with complex problems at all. The defensible inferences are the post-data inferences.⁴⁹

What matters is how sound the inferences are in light of the arguments and evidence presented, not in comparison to every twist and turn of analysis before the author arrived at the final conclusions, or what the author would have thought had the data turned out differently.

Returning to the core concern of mitigating bias when assessing likelihoods, first, recall that inference always requires evaluating likelihood *ratios*, which forces us to ask how well the evidence fits with rival explanations. Second, we reiterate our central point regarding scholarly scrutiny: if despite efforts to follow logical Bayesianism, a scholar nevertheless over-estimates how much the evidence favors an inductively-inspired hypothesis, readers can independently weigh that evidence and critically assess the author’s judgments. Subsequent debate may encourage the author to bring more background information to light that was previously used implicitly, or acknowledge that the evidence is not as strong as previously maintained. In our state-building example, readers might contest our assessment that E_1E_2 very strongly favors H_{LRA} over H_R , perhaps suggesting that this evidence only moderately favors the inductively-inspired hypothesis. Open discussion would then result in greater consensus or at least greater clarity on why scholars interpret the evidence differently.

5.3. Integrity

Concern: We need mechanisms to discourage scholars from choosing procedures after the fact to get the results they want, or manipulating evidence to strengthen results.

The first malpractice—*post-hoc* choice of analytical procedures—is a bigger concern for frequentist inference, which require predefined stochastic data-generation models. Within a Bayesian framework for case-study research, we must make judgments about which hypotheses to consider, how to acquire evidence, and how to interpret that evidence. However, the underlying inferential procedure remains the same: apply probabilistic reasoning to update

⁴⁸ Bowers et al. 2015, 16-17

⁴⁹ Jaynes 2003, 421

beliefs regarding the plausibility of rival hypotheses in light of relevant evidence. Analysis always involves assessing priors, assessing likelihoods, and updating probabilities via Bayes' rule. Unlike frequentist statistics, there is no need to choose among sampling procedures, stopping rules, estimators, tests statistics, or significance levels.

The second malpractice, e.g., deliberately cherry-picking evidence, can certainly occur in qualitative research. However, time-stamping does little to deter such abuses. Any scholar intent on exaggerating results or willing to commit fraud can find ways to do so regardless. Ansell and Samuels make similar observations regarding the related issue of results-blind review—it is always possible to “sweep dirt an author wants no one to see under a different corner of the publishing carpet.”⁵⁰ As a device for signaling integrity, mechanisms like pre-registration or time-stamping risk imposing a substantial burden of time and effort on honest scholars without preventing dishonest scholars from sending the same credibility signals. Recall that the retracted LaCour-Green study was pre-registered, yet evidence of fraud was uncovered not by comparing the published article to the pre-registration plan, but by scrutinizing the article and accompanying dataset.⁵¹

The only viable strategy in our view involves disciplinary norms. First, we need a commitment to truth-seeking and scientific integrity. As Van Evera observed long before DART: “Infusing social science professionals with high standards of honesty is the best solution.”⁵² Second, adjusting publication norms regarding requisite levels of confidence in findings would mitigate incentives for falsely bolstering results. For qualitative research, embracing Bennett and Checkel's dictum, “conclusive process tracing is good, but not all good process tracing is conclusive,”⁵³ would be a major step toward reducing temptations to overstate the case in favor of the author's hypothesis. An associated best practice entails explicitly addressing the pieces of evidence that on their own run most counter to the overall inference; transparency of this type could both encourage critical thinking and signal integrity in a more meaningful way. We recognize that these suggestions are neither panaceas nor quick fixes. But in the long term, rethinking disciplinary norms and practices along these lines and adopting a more a Bayesian perspective could help us better acknowledge and communicate the uncertainty that surrounds our inferences, which is critical for scientific inquiry.

References

- Ansell, Ben, and David Samuels. 2016. “Journal Editors and ‘Results-Free’ Research: A Cautionary Note.” *Comparative Political Studies* 49(13):1809-1815.
- Beach, Derek, and Rasmus Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. University of Michigan Press.
- Bennett, Andrew, and Jeffrey Checkel, eds. 2015. *Process Tracing in the Social Sciences: From Metaphor to Analytic Tool*. New York: Cambridge University Press.
- Bowers, Jake, Jonathan Nagler, John Gerring, Alan Jacobs, Don Green, and Macartan Humphreys. 2015. “A Proposal for a Political Science Registry.” <http://blogs.bu.edu/jgerring/files/2015/09/AproposalforaPoliticalScienceRegistry.pdf>

⁵⁰ Ansell and Samuels 2016, 1810

⁵¹ Broockman et. al 2015

⁵² Van Evera 1997, 46

⁵³ Bennett and Checkel 2015, 30

- Broockman, David, Joshua Kalla, and Peter Aronow. 2015 "Irregularities in LaCour (2014)." https://stanford.edu/~dbroock/broockman_kalla_aronow_lg_irregularities.pdf
- Büthe, Tim, and Alan Jacobs. 2015. "Conclusion: Research Transparency for a Diverse Discipline." *Qualitative and Multimethod Research: Newsletter of the American Political Science Association's QMMR Section* 13(1):50-63.
- Brady, Henry, and David Collier. 2010. *Rethinking Social Inquiry*. Lanham: Rowman and Littlefield.
- Cox, Richard. 1961. *The Algebra of Probable Inference*. Johns Hopkins University Press.
- Fairfield, Tasha, and Andrew Charman. 2017. "Explicit Bayesian Analysis for Process Tracing," *Political Analysis* 25(3):363-380.
- Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishers.
- Gregory, Phil. 2005. *Bayesian Logical Data Analysis for the Physical Science*. Cambridge University Press.
- Humphreys, Macartan, and Alan Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109(4):653-73.
- Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21:1-20.
- Hunter, Douglas. 1984. *Political/Military Applications of Bayesian Analysis*. Boulder: Westview Press.
- Jackman, Simon, and Bruce Western. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88(2):412-423.
- Jacobs, Alan. 2018. "Pre-registration and Results-Free Review in Observational and Qualitative Research." Draft chapter prepared for Colin Elman, John Gerring, and James Mahoney, eds. *The Production of Knowledge*.
- Jaynes, E.T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.
- Jeffrey, Richard. 1983. *The Logic of Decision*. University of Chicago Press.
- Jefferys, William. 2003. "Bayes' Theorem," *Journal of Scientific Exploration* 17(3:537-42).
- _____. 2007. "Bayesians Can Learn from Old Data." <https://repositories.lib.utexas.edu/bitstream/handle/2152/29425/BayesiansOldData.pdf?sequence=1>
- Kapiszewski, Diana, Lauren MacLean, and Benjamin Read. 2015a. *Field Research in Political Science*. Cambridge University Press.
- _____. 2015b. "Reconceptualizing Field Research." Unpublished Manuscript.
- King, Gary, Robert Keohane, and Sidney Verba (KKV). 1994. *Designing Social Inquiry*. Princeton University Press.
- Kurtz, Marcus. 2009. "The Social Foundations of Institutional Order: Reconsidering War and the 'Resource Curse' in Third World State Building." *Politics & Society* 37(4):479-520.
- Laitin, David. 2013. "Fisheries Management." *Political Analysis* 21:42-27.
- Lieberman, Evan. 2016. "Can the Biomedical Research Cycle be a Model for Political Science." *Perspectives on Politics* 14(4):1055-66.
- Loredo, T.J. 1990. "From Laplace to Supernova SN1987A: Bayesian Inference in Astrophysics." In P.F. Fougere, ed., *Maximum Entropy and Bayesian Methods*, The Netherlands: Kluwer Academic Publishers.

- MacKay, David. 2003. *Information Theory, Inference, and Linear Algorithms*. Cambridge University Press.
- Mahoney, James. 2015. "Process Tracing and Historical Explanation." *Security Studies* 24:200-218.
- McKeown, Timothy. 1999. "Case Studies and the Statistical Worldview." *International Organization* 53(1):161-190.
- Monogan, James. 2015. "Research Preregistration in Political Science." *PS: Political Science and Politics* 48(3):425-29.
- Ragin, Charles. 1997. "Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research." *Comparative Social Research* 16:27-42.
- Sivia, D.S., 2006, "Data Analysis—A Dialogue With The Data," in *Advanced Mathematical and Computational Tools in Metrology VII*, P. Ciarlina, E. Filipe, A.B. Forbes, F. Pavese, C. Perruchet and B. Siebert (eds.), World Scientific Publishing Co.:108-118.
- Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Cornell University Press.
- Western, Bruce. 2001. "Bayesian Thinking about Macrosociology." *American Journal of Sociology* 107(2):353-78.
- Yom, Sean. 2015. "From Methodology to Practice: Inductive Iteration in Comparative Research." *Comparative Political Studies* 48(5):616-644.
- _____. 2018. "Analytic Transparency, Radical Honesty, and Strategic Incentives," *PS: Political Science & Politics* 51(2):416-421.

Acknowledgments

The authors thank Andy Bennett, Ruth B. Collier, David Collier, Justin Grimmer, Macartan Humphreys, Alan Jacobs, Jack Levy, James Mahoney, Jason Sharman, Hillel Soifer, and Elisabeth Wood for detailed comments and intellectual engagement. They are also grateful to journal editor Michael Bernhard, Devin Caughey, Christopher Darnton, Steven Goodman, Jacob Hacker, Antoine Maillet, James Mahon, Craig Parsons, Jessica Rich, and Ken Shadlen, as well as seminar participants at CASBS, IQMR, Rutgers, Princeton, Yale, UT Austin, UC Berkeley, and the University of Oregon.