Schaffer, Frederic C. 2015. *Elucidating Social Science Concepts: An Interpretivist Guide*. Routledge.

Scott James C. 1985. *Weapons of the Weak : Everyday Forms of Peasant Resistance*. New Haven: Yale University Press.

Soss, Joe. 2018. "On *Casing* a Study versus Studying a *Case*." *Qualitative and Multi-Method Research*. 16, no.1: 21-27.

Soss, Joe and Vesla Weaver. 2017. "Police Are Our Government: Politics, Political Science, and the Policing of Race-Class Subjugated Communities." *Annual Review of Political Science*. Vol. 20: 565-591

Simmons, Erica, & Nicholas Rush Smith. (Eds.). 2021. *Rethinking Comparison: Innovative Methods for Qualitative Political Inquiry*. Cambridge: Cambridge University Press.

QMMR Qualitative & Multi-Method Research

Book Symposium: Tasha Fairfield and Andrew Charman. 2023. *Social Inquiry and Bayesian Inference*. New York: Cambridge University Press

Qualitative and Multi-Method Research | Fall 2022, Vol. 20.2 | https://doi.org/10.5281/zenodo.8326344

A Seminal Achievement: The First Comprehensive Approach to Formal Bayesian Process Tracing

Andrew Bennett
Georgetown University

I am more excited about the publication of Tasha Fairfield and Andrew Charman's *Social Inquiry and Bayesian Inference: Rethinking Qualitative Research* (2022; hereafter cited in text as SIBI) than I have been about any book for many years. Even for those who prefer to use Bayesian logic informally rather than using explicit priors and likelihood ratios, SIBI greatly clarifies the Bayesian logic that underlies process tracing, and it provides clear guidance for avoiding inferential errors. As Macartan Humphreys once put it to me, Bayesian analysis makes transparent and more reliable the judgments we had to be making anyway to make causal inferences from case studies.

SIBI vaults the discussion of Bayesian process tracing forward on many fronts: how Bayesianism differs from other approaches, how to deal with complications like multiple hypotheses rather than just hypothesis H and its negation (~H or "not H"), the pros and cons of informal and formal Bayesian analysis of evidence from cases, and improvements over existing practical advice on carrying out process tracing. Above all, SIBI makes an enormous contribution by showing that Bayesian logic can in principle be used fully and transparently on every piece of evidence to adjudicate among alternative explanations of a case, even if in practice, as SIBI's authors note, it would be unwieldy to present readers with such a full and formal analysis.

Fairfield and Charman (2022) accomplish these feats while still making SIBI accessible to graduate students and useful for instructors. They provide clear guidelines, numerous exercises, and many worked examples of their approach, relegating he more technical material to appendices. As a result, SIBI is useful both for readers interested in working through all the math and those who prefer simply to understand the intuitions behind Bayesianism and follow the steps required to use its logic in process tracing, whether formally or informally.

In this brief review, I focus on SIBI's contributions on four issues that have often been misunderstood by critics and students (SIBI outlines several of these, and other common misunderstandings, 448-54). These include: 1) the distinction between the logical mutual exclusivity of hypotheses, which Bayesian inference requires, and mutual exclusivity of variables between hypotheses, which Bayesian inference does not require; 2) the number of comparisons among hypotheses vis-à-vis the evidence that are necessary as the number of hypotheses grows; 3) the relevance of evidence to *all* the alternative hypotheses rather than to just one hypothesis or another; and, 4) the value of using the log-odds form of Bayes Theorem.[1] I conclude with four issues that deserve continuing discussion and research: 1) case specific versus population level priors and scope conditions, and ways of generalizing from the process tracing results of individual case studies to wider populations; 2) advice on case selection for small-n research; 3) different ways of estimating priors and likelihood ratios; and 4) the relationship between SIBI and other approaches that use Bayesian logic (particularly Macartan Humphreys and Alan Jacobs's 2023 book, *Integrated Inferences: Causal Models for Qualitative and Mixed Method Research,* on how to use causal models and Bayesian reasoning to integrate qualitative and quantitative research).

While traditional process tracing advice has emphasized the need to consider a wide range of alternative hypotheses, SIBI puts into sharper focus the need for constructing alternative explanations for the outcome of a case that are logically mutually exclusive and also exhaustive (MEE). As Fairfield and Charman point out, if alternative explanations are not mutually exclusive, it makes little sense to ask which provides the best explanation, and it is difficult to think of how one might attach priors and likelihood ratios to overlapping hypotheses (SIBI, 86). Yet as the authors note, the requirement of mutual exclusivity has often been misunderstood as requiring that alternative hypotheses must be monocausal or include only one variable. As they note, "mutual exclusivity of hypotheses is conceptually distinct from exclusivity of their constituent independent variables, causal factors, or mechanisms" (87). Not only can hypotheses include multiple independent variables and still be logically mutually exclusive, they can also include many or all of the same independent variables and be mutually exclusive, so long as they posit different functional relationships among the variables.[2] An internal combustion engine, for example, needs oxygen, fuel, spark, and compression to function. If the engine is not working, it could be that the spark plug is somewhat fouled and the oxygen intake and fuel lines are partly clogged, or it could be a partial malfunction of both the spark plug and the fuel line, or the malfunction could be due to a number of other combinations involving one, two, three, or all four of these same features. These alternative explanations are logically mutually exclusive in that they posit different functional explanations, and they attach to different counterfactuals on what interventions would be necessary for the engine to run smoothly.

It is certainly true that it can be difficult to construct a satisfactory set of mutually exclusive hypotheses, but this is a feature of the complexity of the world and our limited understanding of it, not a consequence of using Bayesian logic. One can always make alternative explanations mutually exclusive by attaching to each of them the claim that it is the most important factor—there can only be one most important factor. It is more useful, however, to construct mutually exclusive hypotheses that have functional differences and that therefore relate more clearly to observable implications that are more likely under one hypothesis than another. One useful starting point for constructing such hypotheses is a typology of theories about causal mechanisms that I have developed. The typology includes twelve families of theories that result from the intersection of four agent-structure relations (agent→agent, agent→structure, structure→agent, and structure→structure) and three categories of explanation that are common in the social sciences (including those that focus on ideas and legitimacy, material power, and transactions costs and institutional efficiency).[3]

The challenge of constructing an exhaustive set of hypotheses, or a set whose probabilities sum to 1, is in some sense the more demanding requirement. As Fairfield and Charman point out, we can never be fully sure we have satisfied this criterion, as it is always possible that an explanation we have not thought of is the best explanation. This is why Bayesians never put 100% certainty on an explanation even if very strong evidence gets them close to 100% confidence. The most we can hope for, the authors note, is "inference to the best *existing* explanation" (SIBI, 84), but they add that we can always add new explanations and reanalyze the evidence in light of the new set of hypotheses; as they note, this is a common practice in science (85).

A second misconception that SIBI puts to rest is the idea that the number of likelihood ratios one must consider, or the comparisons one must make between hypotheses for each piece of evidence, grows combinatorially large as the number of hypotheses increases.[4] In fact, as SIBI demonstrates, it is not necessary to compare every hypothesis to every other hypothesis

1 On the first three of these issues, see Zaks 2021, the response by Bennett, Charman, and Fairfeld 2022, and the rejoinder by Zaks 2022.

2 Fairfield and Charman note a related misconception, which is the idea that alternative explanations must always or mostly make mutually exclusive predictions about evidence. In fact, alternative explanations may make observationally equivalent predictions on many pieces of evidence—they need only make different predictions in at least one actual or possible instance (SIBI, 89).

3 Bennett 2013.. Bennett and Mishkin 2023 adds to this framework theories about intra-agent mechanisms of behavior.

4 Zaks, 2021.

vis-à-vis each piece of evidence. One need only arbitrarily choose one hypothesis and compare the likelihood of evidence under that hypothesis to the likelihood of that same evidence under each of the other hypotheses, and then one has implicitly compared the likelihood of the evidence under all the hypotheses to each other. If we know the likelihood ratio of H1 to H2 and H1 to H3 for evidence E1, then we know the likelihood ratio of H2 to H3 vis-à-vis E1. The analogy I use here is that one need not weigh every item in the grocery store to know their relative weights—we can weigh how many peanuts to a watermelon and how many to a cantaloupe, and then we also know the relative weights of the cantaloupe and the watermelon without ever directly weighing one against the other. Thus, adding a new hypothesis to the existing set of hypotheses requires only one additional comparison for each piece of evidence.

A third mistake that students often make when first learning Bayesian analysis, and one that even some methodologists slip into through poor wording, is the idea that evidence is "on" or "relevant to" or "an implication of" only one hypothesis.[5] SIBI underscores that a critical feature of Bayesian inference is that evidence has some probability under every hypothesis, and it is the *relative* likelihood of the evidence under different hypotheses that determines the probative weight of the evidence.

A fourth issue that SIBI makes admirably clear, but one that nonetheless still causes some confusion among students, is the value of using the log odds form of Bayes Theorem and an associated logarithmic scale, such as the decibel (dB) scale. As Fairfield and Charman point out, using the log odds form of Bayes Theorem greatly simplifies the mathematics of summing up the inferential weights of different pieces of evidence. In addition, our sensory systems for sight, hearing, etc. follow logarithmic scales – our ears can detect small differences in loudness or air pressure between different quiet sounds, but when sounds are already loud, our ears require bigger increments of additional air compression to discern any difference. Fairfield and Charman's suggestion for using the decibel scale to assess the weight of evidence is thus eminently sensible, and they discuss at length (SIBI, 129-36) how to think about and use this scale, as well as providing a table showing equivalent dB and odds ratios (133). Even so, I have found that students require considerable practice to be able to intuitively translate among dB, odds ratios, and percentage probabilities, and practice with a more detailed conversion chart (such as this) can be helpful.

In addition to these and many other contributions, an admirable feature of SIBI is its methodological pragmatism. While I continue to encounter people who think that those of us exploring formal Bayesian process tracing are advocating excessively ambitious uses of the method, to my knowledge *literally no one* has ever advocated that formal Bayesian process tracing should be employed and written up for every piece of evidence from a case study. Fairfield and Charman are careful, both in SIBI and in their earlier work, to acknowledge the limitations of formal Bayesian process tracing and the uncertainties it entails (indeed, as they point out, Bayesian analysis can be thought of as a means to estimate the uncertainty that inevitably remains in any study, not just a method for trying to reduce it). They also point out that it would be incredibly tedious for a reader to wade through a formal analysis of every piece of evidence in a study. I expect that a range of practices is likely to emerge:

- researchers may use Bayesian insights to strengthen informal or traditional process tracing and reduce inferential errors without ever writing up a formal Bayesian analysis of evidence

- researchers might perform formal Bayesian analysis of one or a few pieces of evidence, which they may or may not present to readers in the main text, footnotes, or appendices

- researchers might do formal Bayesian analysis on much or even all of the evidence, but only present the most important parts of this analysis (the pieces of evidence with the greatest inferential weight) to readers, as well as summary conclusions of the analysis

- researchers might do full formal Bayesian analysis of all the evidence in a study, present the most important parts in a publication, and present the rest of the full formal analysis in an online appendix.

I expect the first two of these practices will be the most common, and I would be surprised if a formal Bayesian analysis of all the evidence from a case is ever published in full, even in a book-length project. Nonetheless, demonstrating that a full and formal Bayesian analysis of case study evidence is possible, as SIBI does, is tremendously important. Not only does it clarify the logic of process tracing, it also outlines that logic in a mathematical form that quantitative methodologists and researchers find compelling and legitimate.[6]

I conclude with four issues that deserve further discussion and research. First, generalizing from the results of Bayesian analysis of evidence in one case to a wider population of cases is a complex proposition. SIBI

5 On this point see Bennett, Charman, and Fairfield 2022.

6 I was once the sole qualitative methodologist at a dinner with a half-dozen faculty who teach quantitative methods. One of them asked me, perhaps just to be polite, what was new and different in qualitative methods. When I responded, "formal Bayesian process tracing," the whole group snapped to attention.

devotes chapter five to this issue and offers sensible advice, but I suspect in practice generalization is often more complicated than the examples it discusses. As Fairfield and Charman note, with considerable understatement, in social science "scope conditions are not always explicitly stated from the outset (172)." They also acknowledge that social scientists often use what they call "patchwork hypotheses," or hypotheses that "different causal logics operate in different regions of the overall scope space" (Alex George and I have called these "contingent generalizations"). My default assumption has always been that in social life there are few simple hypotheses with broad scope conditions, so "patchwork hypotheses" are the norm. SIBI outlines procedures for dealing with such hypotheses, but a further complication is that our understanding of scope conditions can change markedly during case study research because as our understanding of the mechanisms in a case change our understanding of their scope conditions often change as well. In addition, it is difficult for scholar to articulate their background knowledge of all the cases in a population, and which pieces of background knowledge they think are important will change as their understanding of mechanisms and their scope conditions change. It is still possible to parse all of this out in Bayesian terms, as SIBI does, but I expect many adjustments are necessary in applying the posteriors on hypotheses from one case study to other cases that we already know are dissimilar in many potentially important respects.

Second, SIBI has a terrific chapter on case selection in small-n research (chap. 12), providing the most comprehensive discussion I have read of all the different approaches that have been proposed. SIBI's argument is that the best criterion for case selection is expected information gain, but that we cannot assess this a priori since we don't know the evidence and likelihood ratios of a case until we gather the evidence. At the same time, the authors maintain, we can expect to learn something from almost any case. Therefore, we should not worry too much about choosing cases that have less (*a priori* unknowable) information gain than other cases we might have chosen (567), and we should be transparent and unapologetic in giving pragmatic rationales for case selection. Still, the authors provide useful Bayesian advice on case selection (567-78): diversity among cases is generally good, similarities across cases can contribute to strong tests, there is no need to avoid cases with multiple plausible causes, and model-conforming cases are good for inferences on mechanisms while deviant cases are good for building or testing higher-level theories. They also sharply critique the concept of most- and least-likely cases.

I concur with these suggestions and insights, but the argument on which I am least certain is the claim that the least/most-likely designation is entirely unworkable, and related, I have not entirely given up on the idea that we can have case-specific priors. It is possible that my somewhat different inclinations from the authors here are simply semantic. What I think of as a "case specific prior" they might call (perhaps more accurately) background information that bears on whether the case fits the scope conditions of a theory. The problem here is that I think it is difficult to articulate all of the background knowledge about both theories and cases that informs scope conditions in sufficient detail that we can treat these scope conditions as binary as SIBI suggests (584, fn 40). Indeed, the authors themselves argue that in trying to assess the scope conditions of a theory, "it makes sense to examine more cases near the boundaries of our scope space (p. 216)," which might be read as implying that our concepts of scope conditions can be probabilistic rather than binary. Or perhaps this is a mis-reading – it comes down to whether we are treating scope conditions as inherently binary, or as probabilistic in the quantum sense, and whether we are accordingly treating uncertainty mostly or only as a reflection of our incomplete understanding of scope conditions (the typical Bayesian view) or as a feature of ontologically probabilistic scope conditions.

Consider a medical example. A doctor might have pretty strong knowledge about some of the scope conditions of theories bearing on the probability that a patient who walks into their office has ovarian cancer. If they are a male, usually a piece of background information that is evident upon first sight, the probability is zero—we could pretty clearly call this a case with known or quickly updated background information that places it outside the scope conditions for any theory of ovarian cancer. But sex is not always biologically binary due to the possibility of hermaphroditism, so there is already some uncertainty for the doctor, whether we are attributing it to possible measurement error on the background conditions or uncertainty on the scope conditions (given the infrequency of hermaphroditism, for example, there may not be adequate research on the incidence of ovarian cancer for hermaphrodites). If the patient is biologically a female (again, with some uncertainty) and the doctor already knows the patient has a mutation on the BRCA1 gene, their probability of ovarian cancer is higher than that of the general population of women. But there are many other attributes of the patient on which either the research or the doctor's knowledge of theories and their scope conditions is fuzzy: age, ethnicity, general health, etc., and their incidences of ovarian cancer. Still, the doctor's general biological knowledge and the (possibly mixed) results of research might allow educated guesses on how these attributes (to some degree instantly updated on seeing the patient) might affect the patient's likelihood of having ovarian cancer, whether we are calling that a case-specific prior or a probabilistic estimate of whether the patient falls into the scope conditions of probabilistic theories about ovarian cancer. I don't think Fairfield and Charman would disagree about the logic of the inferences involved here – it may just be that people typically use the term "case specific prior" for what Fairfield and Charman I expect would call, more accurately, a combination of less-than-certain and incomplete but often quickly updated background knowledge about particular cases together with less than complete or certain knowledge about scope conditions. As my impression is that many people tend to think in terms of "case specific priors," however, it will require ongoing efforts to get them to think more precisely in the terms that SIBI uses.

Also, I would slightly qualify the authors' advice on selecting cases and writing up how we did so. They are logically correct that we need not list all known cases before choosing which ones to study, and that listing the cases *not* chosen does convey salient information on inferences from those that were studied. I would put more emphasis, however, on their pragmatic advice that it is enormously useful to list and do preliminary research on a number of salient cases (SIBI, 569). I also think that listing the cases you almost chose, but did not choose, for process tracing is useful because it can clarify the (often pragmatic reasons) for case selection and pre-empt reviewers from criticizing your case selection because you did process tracing on a particular case they think would have been fruitful.

My critiques here are modest and I agree strongly with almost everything in SIBI's discussion of case selection. Qualitative research will be much improved if researchers and reviewers come to agreement around SIBI's advice on this topic. Even so, given long-standing debates on case selection, it will take considerable discussion to get to consensus around SIBI's advice on case selection criteria, even though that advice in my view is incisive and almost entirely correct.

Third, while SIBI provides excellent advice on estimating priors and likelihood ratios,

this is a topic that deserves more research. One question that deserves experimental work, and one on which Tasha, Theo Milonopoulus, and I have made a (thus far unsuccessful) grant application, is whether crowd-sourced estimates of priors and likelihood ratios are superior to those estimated by individual scholars. This could include several variants of crowdsourcing, including experts, non-experts, individuals estimating in isolation and then aggregating their estimates, groups discussing and then estimating, etc. A key challenge here is that we don't have fully articulated, "objective," and 100% true priors and likelihood ratios against which estimates can be measured. The best approximation might be experiments with estimation by subjects from whom one extremely powerful piece of evidence about a case is withheld, but this would bear only upon whether estimates on the rest of the evidence got close to the "true" explanation, not whether estimated priors or estimated likelihood ratios on any given piece of evidence were accurate.

Finally, I would like to hear more on the authors' views, and those of Alan Jacobs and Macartan Humphreys, on the relationship between SIBI and Humphreys and Jacobs (2023)

on using causal models and Bayesian reasoning to integrate qualitative and quantitative research (Alan Jacobs's contribution to the present symposium is an excellent start on this dialogue). I don't think there are any fundamental disagreements between these books, and they are certainly not redundant. But I'd like to hear more on these authors' views, perhaps in a future symposium in this journal.

In sum, Fairfield and Charman have made an enormous contribution by outlining far more clearly than any prior work how Bayesian logic can be applied in qualitative research. SIBI is both foundational, building on a long tradition of Bayesian analysis across many fields and getting to the root of critical issues, and practical, offering clear and actionable advice for researchers. As a teacher, reviewer, and practitioner of qualitative methods, I am excited to see the ways in which it is already beginning to improve qualitative research and make it more transparent, and I hope and believe that it will have a profound effect on how qualitative research is conducted and how it is viewed by scholars working primarily with quantitative and other methods.

References

Bennett, Andrew. 2013. The mother of all isms: Causal mechanisms and structured pluralism in International Relations theory. *European Journal of International Relations* 19 (3) 459-481.

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

Bennett, Andrew, and Benjamin Mishkin. 2023. Nineteen kinds of theories about mechanisms that every social science graduate student should know. In Harold Kincaid and Jeroen Van Bouwel, eds., *The Oxford Handbook of Philosophy of Political Science*, 154.

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

Humphreys, Macartan, and Alan M. Jacobs. 2023. *Integrated Inferences: Causal Models for Qualitative and Mixed-Method Inference*. New York: Cambridge University Press.

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

Zaks, Sherry. 2022. Return to the Scene of the Crime: Revisiting Process Tracing, Bayesianism, and Murder. *Political Analysis, 30* (2), 306-310.

Qualitative and Multi-Method Research Fall 2022, Vol. 20.2 https://doi.org/10.5281/zenodo.8326449

Leaning In to Analytic Explicitness

Alan M. Jacobs
University of British Columbia

Tasha Fairfield and Andrew Charman's *Social Inquiry and Bayesian Inference* (2022) constitutes a major contribution to the advancement of qualitative methods in our discipline. The volume provides (as far as I am aware) the first extended treatment of Bayesian qualitative inference in the social sciences, covering both the conceptual underpinnings of Bayesianism and a range of issues that arise in its practical implementation. The book's guidance is elaborated with a large number of detailed applications using real data, including re-analyses of the evidence in prominent published works of qualitative political science.

As Fairfield and Charman point out, a Bayesian approach holds the promise—among other virtues—of making qualitative research considerably more analytically explicit,[1] in two related respects. First, carrying out formal Bayesian procedures allows researchers to show exactly how they have made the leap from evidence to inference. Given a stated set of priors over the hypotheses and likelihoods of the evidence under each hypothesis, it becomes straightforward for readers to see where posterior beliefs come from once the evidence is (or is not) observed. Of course, priors and likelihoods must themselves be defended, and readers might disagree about the probabilities assigned by the researcher: explicitness provides no assurance of arriving at the right or a consensual answer. But by surfacing the key premises on which inference is grounded, a formal Bayesian approach makes the analysis far more susceptible to evaluation and critique.

Second, as Fairfield and Charman also make clear, a Bayesian approach provides researchers with a principled way of *aggregating* inferences across multiple pieces of evidence. The problem of aggregating across many pieces of evidence may be modest in situations in which all or nearly all of the evidence points in the same direction. Combining observations becomes much trickier, however, when different pieces of evidence pull in different directions. How certain should we be about a hypothesis if, say, many observations line up in its favor, but a few key pieces of evidence cut against it? Conventional, informal approaches to case-study research will typically struggle with this sort of situation because they tend to lack a principled way of *weighting* observations relative to one another. If researchers are willing quantify their priors and the likelihoods of the evidence, however, formal Bayesianism offers a powerful and transparent mechanism for drawing conclusions from an arbitrarily mixed evidentiary pattern (see pp. 116-117, sec. 3.7).

Of course, no treatment of a method—even one as clear and comprehensive as this book—can fully address all problems or complications that the approach might confront. In the remainder of this essay, I will briefly raise a few issues that I think this book leaves unresolved. I will discuss, in turn, Fairfield and Charman's approach to generalizing from cases; their defense of informalism in the derivation of priors and likelihoods; and their advice on writing up formal Bayesian analyses. Particularly on the last two points, one overall theme of my comments is to suggest that Fairfield and Charman might have leaned even further than they do into Bayesianism's potential to make qualitative inference more analytically transparent and evaluable.

How Do Inferences Travel?

Suppose I have gathered and assessed the evidence from one or a small handful of cases: what can that evidence tell me about other, perhaps similar, cases? Fairfield and Charman (2022) come closest to addressing this question in Chapter 5, where they apply their framework to the qualitative analysis of multiple cases.

1 For an excellent discussion of how different qualitative approaches, including Bayesian analysis and more informal approaches, vary in their "explicitness," see the Qualitative Transparency Deliberation working group report by Kreuzer and Parsons (2018).

In this chapter, they describe an approach in which hypotheses come with *scope conditions* attached to them. The researcher then proceeds to collect evidence from one or more cases that fall within these scope conditions. To update on the hypothesis (with its stated scope conditions), we simply add up the weight of the evidence across the cases examined, arriving at posterior odds ratios for any given pair of rival hypotheses of interest. Fairfield and Charman also consider the auxiliary problem of how to generalize *beyond* the initial scope conditions, but my concern here is with the narrower question of how we learn across cases *within* the original scope conditions.

Fairfield and Charman work through their approach, in part, with an application to Dan Slater's research on democratic mobilization in authoritarian Southeast Asia, considering three hypotheses: one focused on the role of autonomous communal elites in fostering mobilization, a second positing economic decline as the central factor, and a third centered on stolen elections. Fairfield and Charman articulate each hypothesis with the region of Southeast Asia as an explicit scope condition. They then use Slater's evidence from two cases—the Philippines and Vietnam—to update beliefs over the three hypotheses. The weight of each piece of evidence observed, regardless of the case from which it is drawn, is simply added together to yield the relevant posterior odds ratios (over any two of the three hypotheses that we might want to compare).

If I have understood the approach here correctly, because we are always updating on the hypotheses—and because these hypotheses are framed in terms of some *set* of cases, such as autocratic countries in Southeast Asia—the posterior beliefs that we generate are always understood to apply to *all* cases that fit the stated scope conditions. In Fairfield and Charman's reanalysis of Slater's data, the weight of the evidence in Vietnam and the Philippines overwhelmingly favors the communal elites hypothesis over the economic decline and stolen elections hypothesis. On my reading of Fairfield and Charman's approach to generalization, this means that we now have much greater relative confidence in the communal elites hypothesis as it applies to *all autocracies in Southeast Asia*. We should now believe communal elites to be the overwhelmingly likely cause of any mobilization that we observe in, say, autocratic Thailand or Malaysia because of the evidence observed in Vietnam and the Philippines.

It certainly seems intuitive that what we observe in Vietnam and the Philippines should affect our beliefs about other cases that share similarities to these two. But what seems odd to me is that there does not seem to be any mechanism here for distinguishing our posterior beliefs about those cases from which we *have* observed evidence from those cases from which we have *not* observed evidence. In other words, Fairfield and Charman do not appear to build in a role for uncertainty about the degree to which conclusions travel across the domain of theoretical interest. Lesson-drawing across cases seems to be automatic, the problem of generalization apparently assumed away by the declaration of a scope condition.

I am further perplexed by Fairfield and Charman's insistence that we should aggregate the weights of the evidence in exactly the same way regardless of whether the evidence all comes from within a single case or is spread across multiple cases. Either way, as long as the evidence derives from within the stated scope condition, we are simply updating on the hypothesis. Thus, for instance, there is no distinction to be made between observing, say, three highly probative, independent pieces of evidence in favor of the communal elites hypothesis (relative to its rivals) within a single case, on the one hand, and observing those same three highly probative pieces spread across three separate cases, on the other hand. I would have thought that, unless we have strong prior beliefs about the homogeneity of cases within the scope condition, we would want to shift our beliefs more strongly in favor of the communal-elites theory under the second scenario (evidence spread across cases) than under the first (evidence all within one case).

A simple thought experiment makes especially clear what Is problematic about automatic generalization across a scope-condition-defined domain. Suppose that instead of framing the Slater hypotheses as applying to autocracies in Southeast Asia, we started by framing the hypotheses as applying to *all* autocracies (and there is nothing intrinsic to the three hypotheses that makes this implausible). Despite having dramatically expanded the hypotheses' scope, there is nothing I can see in Fairfield and Charman's approach that changes how we would update on these much more general hypotheses from, say, evidence on Vietnam and the Philippines.

Defending Fairfield and Charman's approach to generalization, at least as articulated in the book, would seem to require defending very strong assumptions of exchangeability or homogeneity across cases within a given set of scope conditions. Such assumptions will not usually be tenable in social scientific applications. An alternative approach—one that would still be broadly consistent with Fairfield and Charman's framework, I think—would involve building the researcher's beliefs about heterogeneity directly into the likelihoods of the evidence, thus allowing these beliefs to condition the portability of findings across cases. Doing so would still allow for generalization and cross-case learning, but in more sensible ways. It would have us update *more* strongly about the cause of mobilization in Thailand from evidence drawn from Thailand than from evidence drawn from Vietnam. It would generate sensibly weaker generalizations across domains that the researcher believes to be highly heterogeneous than across those that are believed to be more homogeneous. And it would take into account how the evidence is distributed across the domain, including, for instance, whether we are observing similar patterns of evidence across cases that we were not *a priori* confident would exhibit similar causal relationships or mechanisms.

Where Do Priors and Likelihoods Come From?

Fairfield and Charman's approach formalizes inference starting from the point at which the researcher states prior beliefs about the hypotheses and (relative) likelihoods of the evidence under the hypotheses. How one derives priors and likelihoods, however, is left almost completely informal. I refer the reader to Chapter 3 for Fairfield and Charman's interesting discussion of how researchers should "inhabit the world of each hypothesis" (p. 105) to informally reason their way to their likelihoods.

It is surely impossible to formalize all aspects of any research process, and I have no quibble with Fairfield and Charman's decision to limit their own formalization to the process of inference from evidence, given a set of priors and likelihoods. What I would take issue with, however, is Fairfield and Charman's defense of this choice as reflecting fundamental limits of formalization.

One way of formalizing the generation of priors and likelihoods would be to begin with a formal *theory* of the causal processes operating in the domain of interest, perhaps expressed as a probabilistic causal model (Pearl 2009). In brief, by positing prior probability distributions over exogenous conditions, one can then use the model to derive priors on the probability of alternative causal effects or processes unfolding and about the likelihood of observing a given piece of evidence under the operation of alternative effects or processes.[2]

In Chapter 9, Fairfield and Charman (2022) argue persuasively that in most social scientific contexts, as opposed to some natural-scientific domains, we are unlikely to be able to arrive at *objective* groundings of our likelihoods. In the "hard" sciences, they point out, "strong underlying theory and well-understood error models for the measurement apparatus" (441) sometimes yield unambiguous likelihood functions with strong empirical groundings. These are conditions that rarely prevail in social scientific research situations, meaning that our likelihoods will always contain a large element of subjectivity. All of this I find persuasive.

What is not obvious to me, however, is how or why the *subjectivity* of likelihoods in the social sciences speaks particularly in favor of *informalism* in the derivation of likelihoods. It is not exactly clear from the text how Fairfield and Charman see the relationship between objectivity and formalization, but they seem to elide the two concepts in arguing against formalized theories as a source of likelihoods, writing:

> We can aim to formalize theories as mathematical models in order to make them more precise, but this approach may give only a veneer of objectivity, in that the model will have to be parameterized, and then further theories and/or prior probability distributions will be needed to inform the values of those parameters, which simply pushes the subjectivity back deeper into the model. (2022, 442)

To critique the use of a model as providing "only a veneer of objectivity" is to miss a couple of the key functions of a model, even of a model built on purely subjective assumptions. For one thing, writing down a model representing the researcher's beliefs about how the world works, and from which the likelihoods are then derived, makes explicit elements of the analysis that will otherwise remain implicit. The model may represent a purely subjective set of beliefs, and thus everything that flows from the model will necessarily be model-dependent. But the formalization itself makes clear to the reader exactly what those underlying beliefs are and how they lead to the posited likelihoods—in turn, exposing those beliefs to critical evaluation. In addition, formally deriving priors and likelihoods from a single underlying model forces internal consistency among the inputs to Bayesian analysis, in a way that informal derivation is unlikely to do.

In other words, perhaps differently from Fairfield and Charman, I understand the limits to objectivity and the merits of formalization to be quite distinct issues. To my mind, building Bayesian inference atop formalized theories does not push problems deeper into the analysis; rather, it *extends* the benefits of analytic explicitness deeper into the process of scientific reasoning.

How to Write up Qualitative Bayes?

Whatever the benefits of formalization, formalizing inference undeniably involves tradeoffs. For qualitative

2 Macartan Humphreys and I present a causal-model-based approach to Bayesian inference in a new book (Humphreys and Jacobs 2023). My point here, however, is not about the virtues of any particular approach, but about the general idea of deriving priors and likelihoods from formalized theory.

researchers, one of the steepest of these tradeoffs involves how the empirical evidence is presented.

Qualitative researchers typically deploy a narrative structure in the empirical presentation of case evidence. Narrative can provide a particularly clear way of conveying how different case observations and events are temporally and logically connected. A narrative presentation provides the reader with a contextualized, textured, and relatively holistic understanding of the case and the multiple processes unfolding within it. Moreover, a well-written narrative can be interesting and enjoyable to read.

Bayesian inference, to put it mildly, does not readily lend itself to narrative structure. In Bayesianism qualitative analysis, the holism of a case gives way to the consideration of individual pieces of evidence and their (possibly joint) likelihoods under the rival hypotheses. While it is eminently feasible in Bayesian reasoning to take account of context, temporality, and overall patterns in the evidence, narrative per se is an awkward fit with formal Bayesianism. There is thus a risk that, in adopting a Bayesian approach to qualitative inference and reaping the gains of analytical explicitness, we lose some of the benefits of more conventional modes of qualitative research presentation.

Fairfield and Charman have a proposal for squaring this circle. They recommend that authors start with the story and then go Bayesian:

> Begin with a narrative that describes, interprets, and explains the bulk of the evidence from the perspective of the hypothesis that we consider most plausible. We then proceed to consider rival hypotheses, at which point we can employ either heuristic or explicit Bayesian analysis to evaluate how strongly the evidence supports our inference....If there are some pieces of evidence that fit poorly with the narrative account (e.g., they seem fluky or inconsistent with the author's argument), these can be deferred for explicit consideration in the subsequent Bayesian hypothesis comparison. (2022, 326)

This proposal seems, on its face, to offer the best of both worlds. Those readers who prefer to consume their cases whole will get a narrative; those who prize analytical explicitness will get their priors and odds-likelihood ratios; and the Bayesian analysis is itself helpfully contextualized.

I suspect, however, that the workability of this both-and approach hinges on the researcher's uncovering a rather tidy alignment of the evidence. If the evidence largely lines up in favor of a single hypothesis—as in many of the applied illustrations in the book—then it seems quite straightforward to construct a clear narrative that "describes, interprets, and explains the bulk of the evidence from the perspective of" that hypothesis.

Yet the data are often less cooperative than that: we often end up with a collection of observations pointing in different directions. By this, I do not simply mean that we often find evidence that multiple factors helped shaped an outcome; that is a kind of complexity that can be fairly readily captured in narrative form. What I mean is that we often find a good deal of evidence *supportive* of the claim that factor X mattered to an outcome, together with a good deal of evidence *undermining* the claim that X mattered to the outcome. This is an evidentiary situation that is going to be a much poorer fit with narrative presentation, as there is then no dominant theoretical logic on which to lean in organizing the story. It seems a fairly tall order to construct a story of how things unfolded within a case that is clear and readable, on the one hand, but also faithful to the empirical *uncertainty* about what happened, on the other hand.

Meanwhile, as I noted at the outset, this is the kind of situation to which formal Bayesianism is ideally suited. The problem of evidentiary cacophony is a trivial one from a Bayesian perspective. When we apply the Bayesian apparatus, supporting pieces of evidence shift our beliefs in favor of a given hypothesis relative to its rivals; undermining pieces of evidence shift our beliefs away from that hypothesis; and all shifts are weighted by likelihood ratios indicating how much more or less expected the evidence is under the hypothesis than under its rivals.[3]

My concern is that an approach to the writeup that foregrounds a narrative might only be well suited to situations in which the evidence "cooperates." Or, worse, that it might tend to yield presentations that convey more confidence in the "most plausible" hypothesis than the evidence itself justifies.

To be clear, I do not have in mind a better way of squaring this presentational circle. My main point is to suggest that the trade-off between narrative structure and Bayesian logic is a steeper one than Fairfield and Charman's proposal implies; I suspect that researchers will generally have to choose which they want to prioritize. But I could well be wrong. As Fairfield and Charman's readers begin to craft their own Bayesian case studies, we will likely see much experimentation with presentational form, perhaps giving rise to inventive syntheses between narrative coherence and analytic explicitness.

3 We can, of course, also take dependencies among observations into account in the likelihood function.

References

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

Humphreys, Macartan, and Alan M. Jacobs. 2023. *Integrated Inferences: Causal Models for Qualitative and Mixed-Method Inference*. New York: Cambridge University Press.

Kreuzer, Marcus, and Craig Parsons. 2018. "Epistemological and Ontological Priors: Varieties of Explicitness and Research Integrity." Final Report of the Qualitative Transparency Deliberations Working Group I.1, Subgroup A. August 2018. https://static.cambridge.org/content/id/urn:cambridge.org:id:article:S1537592720001164/resource/name/S1537592720001164sup002.pdf

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2nd edition. New York: Cambridge University Press.

Qualitative and Multi-Method Research | Fall 2022, Vol. 20.2 | https://doi.org/10.5281/zenodo.8326432

Social Inquiry and Bayesian Inference: An "Objective" Vision for Mixed Methods Research?

Sirus Bouchat
Northwestern University

Social Inquiry and Bayesian Inference takes as its premise the idea that Bayesian inference has the power to redefine methodology in political science. Putting itself in the company of works like Rethinking Social Inquiry (Brady and Collier 2010), Fairfield and Charman (2022) position the book as an intervention into the reified divide between qualitative and quantitative research, seeking to elevate Bayesian inference as the unifying framework through which to reposition qualitative research on par with quantitative approaches. The specter of "subjectivity," however, haunts the project throughout, both limiting its capacity to achieve its goals of defining a unifying framework for social scientific analysis, and leaving fundamental questions about research best practices in a Bayesian approach largely unaddressed.

The comprehensive scope of Fairfield and Charman's book reflects its ambitious aim to provide a detailed accounting of how researchers should rigorously specify and evaluate social scientific hypotheses regarding (qualitative) data using Bayesian frameworks. Much of the discourse advocating for Bayesian approaches in social science remains bisected. Quantitative approaches to integrating Bayesian methods into social science research practice range from the technical (e.g., BDA3) to the informal or colloquial. Recent works like Humphreys and Jacobs (2023) increasingly leverage Bayesian reasoning to tackle ongoing challenges across quantitative and qualitative work, such as fundamental questions of causal inference.

Unlike article-length treatments, Social Inquiry and Bayesian Inference has the breadth to provide thorough descriptions of Bayesian tools and paradigms alongside illustrative examples and exercises that make it a particularly powerful teaching tool. Even so, its expansive mandate for engaging qualitative data with Bayesian methods leaves dialogue with quantitative Bayesian approaches largely implicit, or indirectly reflected in sections targeting mixed methodology. From a qualitative research perspective, this book clearly addresses a need for detailed and practical guidance on implementing research within Bayesian logics; from a quantitative perspective, this project misses opportunities for sites of linkage in part because of the conception of "mixed methods" research it invests in. Specifically, as I discuss further below, the limited discussion of prior construction and halting directives around contending with prior probabilities themselves reflects the book's staunch defense of logical and objective Bayesianism—a stance that both limits its ability to champion truly mixed methodology while also creating a perplexing tension with the goal to better integrate Bayesian methods with qualitative approaches.

Mixed Methodology: Unified Inference?

Social Inquiry and Bayesian Inference traverses familiar ground in the space given to articulating differences between Bayesian and frequentist statistical paradigms, noting the fragility of frequentist approaches to any interference in research design as well as the intractability of frequentist interpretation for many questions social scientists would like to ask. By contrast, argue Fairfield and Charman, Bayesian approaches are much more flexible, enable much more nimble use of data, and allow researchers to present their findings in more digestible formats—for example, using credible intervals that have the exact interpretation (e.g., "an *x%* chance of an event occurring") that students are repeatedly cautioned against offering for traditional frequentist confidence intervals.

This fervor for Bayesian methodology in contrast to frequentist approaches lends itself, then, to contending that (logical) Bayesianism provides a unifying bridge among traditionally dichotomized qualitative and quantitative research approaches. Fairfield and Charman (2022) suggest that the unifying quality that Bayes contributes is its inferential framework, noting that the process of updating prior probabilities holds irrespective of whether our data are qualitative or quantitative (382). To contend, though, that Bayesian inference has the potential to create parity across at least the qualitative/quantitative research methods dichotomy requires that the pivotal factor creating hierarchies or status and prestige differences between qualitative and quantitative research is differences of *inference*.

To the extent that this distinction still meaningfully exists, or that an overinvestment in the idea that quantitative methods or data hold higher status remains prevalent, I would argue that charges against qualitative research coming from quantitative scholarship more often pertain to threats to inference at the level of data selection (and the informativeness or bias of those data), than any flaw in the inferential capacity of qualitative *methodology* per se. That is, a concern about a qualitative project drawing inferences based on paired case studies or ethnographic field work might lie in the case selection criteria, or in the ability of ethnographic observation and analysis to truly capture the social or political phenomena of interest to the research question. These issues have less to do with the capacity of case study analysis or ethnographic field research to draw valid inferences based on their data, but rather are concerns about the broader normative or epistemological project of research: how much should we aim for generalizability? How valid (in the substantive sense) are studies that do not identify causation?

Concerns about data quality of bias certainly *threaten* inference, but they are not critiques of inferential process. This matters for the argument Fairfield and Charman lay out for Bayesianism as a unifying paradigm of mixed methods research. While Bayesian updating itself can be leveraged in both qualitative and quantitative domains, it does not at all resolve (and perhaps in fact heightens) concerns about what qualifies as *good data*. The unifying principle of Bayesian analysis, as articulated by Fairfield and Charman—"apply Bayes' rule to update prior odds by evaluating likelihood ratios," (2022, 383)—addresses a higher order concern about having coherent reasoning practices across research designs, but when the leverage you gain to address your research question precisely comes from updating priors with respect to *data*, Bayesianism does not have any inherent capacity to resolve the qualitative vs. quantitative divide that resides primarily in concerns about the validity of the data themselves.

Fairfield and Charman come close to acknowledging this challenge later in the book, noting that "there is no clear procedure for translating complex, narrative-based, qualitative information into precise probability statements" (2022, 441–42), and subjectivity—which they seem to use interchangeably with "arbitrariness," although I disagree—likely arises as a result. Likewise, they acknowledge translating qualitative data to quantitative forms of measurement can induce noise that even careful analysis cannot undo. A truly "mixed methods" project would treat *evidence* derived qualitatively as equal with that measured and collected quantitatively, but the insistence throughout the book on objective Bayesian analysis leads directly to a maligning of subjective measurement or assessment as "arbitrary" at best (erroneous at worst). This distinction all but guarantees that qualitative scholarship remains subject to dismissal or denigration based on its *measurement* strategies or data collection enterprise, and without resolving this distinction, no amount of Bayesian inference can truly unify the epistemological divide.

Chasing Objectivity

Throughout the book, Fairfield and Charman reinforce their allegiance to logical Bayesianism and objective Bayesianism, arguing that these paradigms are the only appropriate and consistent frameworks through which to approach data. In Chapter 9, for example, the authors reify the distinction between qualitative and quantitative research in part by appealing to disciplinary differences in the "hard" science relative to the social sciences: social sciences, they note, "study far more complex and inherently noisier systems" (2022, 441), but rather than leveraging that insight to question the fundamental construction of knowledge and knowledge-generating processes even in the more "objective" hard sciences, they reassert the need to conform as much as possible to objective aims, measurement, and likelihood specification. This defense of objectivity throughout the book, to my mind, limits the possibilities both for truly "mixed methods" research and, puzzlingly, for qualitative research in the social sciences—ostensibly at odds with the book's main goal. Nowhere is this tension, and its implications for the practical application of the approaches detailed in the book, more evident than in the discussion (or lack thereof) of specifying prior probabilities.

Practical Advice about Priors

In contrast to the attention given to specifying and evaluating hypotheses throughout *Social Inquiry and Bayesian Inference*, specifying priors receives relatively little coverage. Priors play an interesting if vexing role in the book as a whole: while heralded as a critically important component of Bayesian analysis and championed as a distinct advantage over frequentist frameworks (e.g., regarding incorporating information from engineering reports in evaluating spacecraft reliability for Mars missions; Fairfield and Charman 2022, 377), priors are also a site of concern about undue influence, subjectivity, and bias.

The detailed guidance and options presented throughout the book for applying Bayesian frameworks (e.g., 118, table 3.1) only make sense conditional on the establishment of prior probabilities, yet *how* precisely a prospective Bayesian researcher should do this is left as an exercise to the reader. That is, although ostensibly the authors allow for priors arising from an informed position (118, table 3.1, option a), appropriately formulating such a prior is not discussed. Per objective Bayes, defining priors over rival hypotheses proceeds from a position of ignorance, and throughout the book this type of prior appears to be the favored solution (either by utilizing a variety of priors somewhat agnostically or by specifying explicitly indifferent priors). Indeed, Fairfield and Charman raise concerns that priors should not be polluted by knowledge of the research design, hypotheses, or evidence when attempting to incorporate "background information."

For both explicit Bayesian analysis and heuristic application of Bayesian logic, they argue that "carefully discussing the strengths and weaknesses of rival explanations based on existing literature" is the obvious first step in formulating and justifying prior selection (2022, 491), but nevertheless seem preoccupied with the idea that analysis might be "sloppy" or involve ad hoc speculation, and secondarily that priors arise post hoc from evidence (492). Rather than detailing procedures for systematically devising sound priors given a review of literature or extensive expertise in a subject matter domain, though, the advice hews toward equal/ignorance priors—a position meant to reflect impartiality and objectivity, but one that instead problematically reflects a direct assertion of ignorance where none truly exists. The overemphasis on avoiding biased or subjective priors further seems misplaced given the authors' acknowledgment that adjudicating multiple priors is a possible option; specifying disagreeable, unreasonable, or biased priors is not inherently problematic, so long as a clear, scientific, and transparent process for re-evaluation is possible.

Notably, Fairfield and Charman (2022) do not only prefer an objective approach on practical grounds, but rather directly position themselves in opposition to subjective Bayesianism and rigorous attempts to instantiate informative priors. They write:

> [Others] might advocate using priors that reflect the collective knowledge or current state of consensus among a relevant community of scholars. While much has been written about eliciting prior probabilities and pooling expert opinion, our logical Bayesian approach is intended to reflect the rational beliefs of the scholar conducting the research. Rather than adopting other experts' probabilities as our own, or averaging priors across multiple scholars, we should conduct our own analysis, while of course drawing on evidence supplied by previous research. In our view, consensus building can best take place subsequently, through collective debate and scrutiny of our work, whereas when assigning priors, authors can and should draw on their own specific background knowledge, which may not be shared by other scholars. (98)

Ceding this ground explicitly weakens the vision of generating a unified approach to mixed methods research, both because it undermines precisely the types of knowledge and expertise qualitative scholars are likely to have (i.e., nuanced perspectives drawn comprehensively from across sources) and because it narrows the scope of research to focus on internal consistency at expense of the broader scientific project of knowledge.

Bayes and the Project of Scientific Knowledge

The visionary aim of *Social Inquiry and Bayesian Inference* to provide a unifying framework for social scientific research is not met with a macro perspective or broader scope for how Bayesian approaches can inform the evolution of scientific knowledge, and specifically how studies using these approaches can build on one another. The effort and attention to detailing how researchers should iterate within their own projects without compromising scientific integrity (e.g., chap. 10) is admirable, but the concern about polluting specified priors with biased information (e.g., 97–98) creates a gap in the specific guidance offered for how researchers should engage prior literature. Fairfield and Charman take for granted that researchers do literature reviews carefully (or should), and that readers will attentively correct specious priors or analyses, but absent concrete direction for incorporating previous research into prior probabilities, the book's detailed micro perspective on rigorous Bayesian inference loses its macro counterpart: a theory of knowledge-building in the social sciences.

Even with its ambitious aims for integrating social science research under a Bayesian umbrella, and its thorough exposition of how Bayesian logic can apply to qualitative data, *Social Inquiry and Bayesian Inference* leaves unaddressed how this (objective) Bayesian approach could integrate research over time, and particularly across disciplines. The authors encourage skepticism, in fact, of research that may reflect "varying degrees of subjectivity in evaluation of likelihood ratios," and directly acknowledge that this "limits what we can reasonably expect in practice" when formalizing Bayesian inferences (2022, 444). The project's dedication to "objectivity" throughout is a particular disservice to the nuance of qualitative scholarship, which does not lack in its scientific value by leveraging data or insights that defy easy quantification, but which nevertheless remains in the shadow of quantitative claims of superiority via "objectivity." Moreover, without clearly delineating how to specify pristine priors, unencumbered by external information and not overly influenced by researcher beliefs, the vision Fairfield and Charman provide for social science research remains insulated and isolated—disconnected from a narrative of how social science research can progress, and knowledge can accumulate.

References

Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* 2nd edition. Lanham: Rowman & Littlefield.

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

Humphreys, Macartan, and Alan M. Jacobs. 2023. *Integrated Inferences: Causal Models for Qualitative and Mixed-Method Inference.* New York: Cambridge University Press.

Qualitative and Multi-Method Research Fall 2022, Vol. 20.2 https://doi.org/10.5281/zenodo.8326462

Bayesian Challenges to Conventional Wisdom and Practice?

Hillel David Soifer
Temple University

In *Social Inquiry and Bayesian Inference*, Tasha Fairfield and Andrew Charman (Fairfield and Charman 2022) seek to provide the most comprehensive foundation for qualitative research in political science by grounding it in the fundamentals of logical Bayesianism. In previously published articles (Fairfield and Charman 2017, 2019) the authors have focused on methods for identifying and evaluating evidence for within-case analysis. But the logical Bayesian approach underpins guidance for a much wider range of research tasks in both qualitative research and beyond. And it is in these areas that the book (hereafter cited in text as SIBI) is especially powerful in breaking new ground.

In this commentary, I engage with three elements of the approach in SIBI in order to think about how the book might be received and read. I begin with their overall project of developing a unifying logic of inference. Second, I reflect on how process tracing is presented in SIBI, since it is here that I expect the book will be most controversial. Third, I highlight some other, more meso-level, ways in which the book challenges the utility of existing research practices and pushes us toward what seem to me more fruitful and practical research design. In my view, these are three salutary challenges to the existing conventional wisdom in the QMMR community. Even if not all readers are persuaded by the case that Fairfield and Charman outline, there is significant value in engaging with the positions that this book outlines.

A Unifying Logic of Inference

I want to focus first on the book's overall orientation to research. Here, Fairfield and Charman are explicit—they believe that logical Bayesianism provides a logic of inference, or more precisely <u>the</u> single logic of inference that unifies all research that seeks to advance causal implications. This is a sharp and explicit pushback against what seems to have become conventional wisdom in the QMMR community—that there are distinct logics of inference, if not even broader differences, between qualitative and quantitative research. Against the view that there are distinct logics of inference (Goertz 2017) or even distinct "cultures" underlying qualitative and quantitative research (Goertz and Mahoney 2012), Fairfield and Charman argue that the logical Bayesian framework accommodates all kinds of data, and treats it all similarly in making and evaluating inferences.

Arguably, this is the boldest and most far-reaching attempt to assert a single logic of evidence underlying all (social science) research that the political science methods community has seen in thirty years. One reading of SIBI, and I don't at all intend this to be uncharitable, is that it is a Bayesian version of KKV (King, Keohane, and Verba 1994). After all, King, Keohane, and Verba argued that there is a single logic of inference that underlies all forms of social inquiry, and that differences among types of data were no more than cosmetic. Of course, as is well known to readers of this publication, KKV was not well received among scholars oriented to qualitative research because it tried to subsume qualitative work into a broadly quantitative paradigm. We might ask, then, about the place of qualitative research in SIBI. Are qualitative scholars going to have a parallel reaction and feel taken aback because Fairfield and Charman subsume their work into a broader paradigm of inference that washes away the unique features or nature of qualitative social science?

Here, I confess that after several readings of the manuscript, I have come to sense a tension in how SIBI conceives of qualitative research. One version of the book's approach is a purely practical one: any data, whether qualitative or quantitative, single-case or cross-case, that is informative as we seek to arbitrate among hypotheses is useful, so we should be qualitative researchers when we find useful data that is qualitative. Perhaps instead of terming this view of research a practical one, we could describe it as omnivorous—SIBI argues that we should consume and integrate into our research any data—of any kind—that is useful.

But I think that at times SIBI evinces hints that the authors have commitments to particular features of a logic of inference that falls closer to the qualitative "culture" described by Goertz and Mahoney (2012). I see a commitment to qualitative research in its own right entering into the presentation through the way SIBI discuss causation itself. For example, Fairfield and Charman write that "a well-specified explanatory hypothesis should generally include some sort of causal mechanism" (SIBI, 80). This assertion is likely not controversial for the typical reader of *QMMR*, though I return below to the question of how Fairfield and Charman approach process tracing. On the other hand, this claim is certainly not fully consistent with some approaches to thinking about causation found in (certain segments of) quantitative research. In other words, SIBI seems grounded in a fundamentally qualitative tradition of how causation should be conceptualized. But this claim that good explanation "should generally include" causal mechanism is not grounded by the authors in the foundations of logical Bayesianism, and indeed it is not justified at all. And much of the book's guidance rests heavily on this claim that causal mechanisms make hypotheses better. I wonder, then, whether much of the attempt to unify qualitative and quantitative methods found in Part III of SIBI will be seen by certain communities of quantitative scholars in a way not unlike how the QMMR community saw KKV—as an attempt to assert a logic of inference that subsumed their research into a paradigm they saw as resting on foundations incompatible with their research practices. More broadly, I expect that many qualitative researchers will be pushed by SIBI to reconsider their orientation toward quantitative research, and towards the question of whether and how distinct research methods can be combined, and knowledge can cumulate across multiple, incommensurate kinds of evidence.

Process Tracing

A related issue, of course, is how Fairfield and Charman think about within-case analysis. Here, I turn from the broad orientation of the book toward more specific research practices. While SIBI is likely to provoke strong reactions from a variety of research communities, this is an issue on which it is especially provocative. Other scholars (Bennett and Checkel 2015; Humphreys and Jacobs 2015; Mahoney 2021) have grounded within-case analysis in a framework of Bayesian updating; that is not provocative in and of itself. Nor is the application of formal Bayesian analysis in my view the novel and notable analytical move that SIBI makes. Instead, SIBI takes a clear and controversial position about what makes within-case analysis informative. Against a robust body of scholarship (Beach and Pedersen 2019) that sees within-case analysis as informative only when it traces steps in the causal process, Fairfield and Charman argue (SIBI, 405ff) that <u>any</u> information that arbitrates among hypotheses is informative. As they write: "the notion that inference entails simply tracing causal mechanisms is a narrow understanding of what constitutes evidence."

A Bayesian logic of inference, then, provides a justification for resolving a debate about the nature of process-tracing in favor of a broader and more eclectic approach to within-case analysis that is not oriented toward causal process alone. I suspect that on this issue, SIBI will face an uphill battle in persuading those committed to the alternative view to abandon their stance. But while previous scholarship that takes this more eclectic view has too often only done so implicitly rather than explicitly justifying this broader view of within-case analysis, Fairfield and Charman make the debate explicit in a salutary way.

Mechanisms Redux

Note, however, that the position here of decentering causal mechanisms in favor of a broader-tend approach to within-case analysis is to some extent in tension with the mechanistic view of causation that (I suggested above) serves to ground the overall project of SIBI. In trying to resolve this for myself, I've come to think that rather than arguing for a mechanistic view of causation in which causal mechanisms are the *sine qua non* of making good causal claims, Fairfield & Charman may instead see causal mechanisms as one sufficient but not necessary way in which scholars can elaborate hypotheses more precisely. Since, as SIBI argues, precise and detailed hypotheses facilitate the use of evidence that arbitrates among them, causal mechanisms are one way that scholars can improve their inferences.

This view, of course, resonates quite strongly with the emphasis in KKV on maximizing the observable implications of hypotheses, acknowledging (as KKV do) that much evidence about causal mechanism is likely to be qualitative. To return to the issue raised at the start, I think there's more to be done to pin down exactly the place of the "mainstream" qualitative research worldview and its emphasis on mechanistic causation in SIBI. If the past few paragraphs here are accurate, they suggest that a certain set of qualitative scholars may see an insufficiently mechanistic view of causation in SIBI and find themselves wary of being subsumed into its unified logic of inference. Just as I argued above, however, I think that by pushing these tensions into the open, and by taking such a clear and well-grounded position on them, SIBI will push scholars to articulate their responses in ways that will move these debates forward in fruitful ways.

Existing Research Practices

In addition to these broader issues, SIBI is likely to provoke and persuade on the more micro-level of research practices. One is the approach to case selection, discussed in the most sustained way in Chapter 12. Here, too, SIBI takes on a robust tradition in qualitative methods scholarship, arguing against many algorithmic practices of case selection in favor of a more practical set of guidelines. Second is the use of all evidence for all hypotheses. This entails among other things a move away from dismissing alternative explanations in a perfunctory fashion in a theory chapter or via claims of controlled comparison toward systematic and thoughtful engagement with alternatives.

There are of course many other points in the book that are valuable touchstones for scholars and worthy of discussion. But these two represent points on which the book raises challenges for standard practices in qualitative research and grounds those new approaches in principles of logical Bayesianism in an especially clear and sustained way. I expect that these are areas in which SIBI will influence research practice in especially far-reaching ways: I, for one, have already been advising students to take both of these practices on board in designing and carrying out their research, and I look forward to assessing the extent to which others do as well.

In closing, it should be clear that SIBI is poised to be transformative at three levels. Working backwards through this essay, we can see that it has the potential to change existing research practices, to fundamentally reshape debates about the nature of process-tracing, and to invite new conversations about whether and how social scientific inference can be unified under a single logic. That, to put it mildly, is no small accomplishment: many of us will never write anything that shapes the way so many scholars think about and carry out their work. But scholars may use this book to justify the positions they take on these three levels without fully taking on board its underlying framework of logical Bayesianism. To what extent will the authors be satisfied in moving us a bit closer to practices consistent with Bayesianism even if we don't take on board the underlying logic? Will Fairfield and Charman be content if we all act a bit more Bayesian, or is the goal here to convert us into Bayesians? I look forward to hearing their response today, and to continuing what has already been a very fruitful conversation and learning experience over the years to come.

References

Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process Tracing Methods*. Ann Arbor: University of Michigan Press.

Bennett, Andrew, and Jeffrey T. Checkel, eds. 2015. *Process Tracing: From Metaphor to Analytical Tool.* Cambridge: Cambridge University Press.

Fairfield, Tasha, and Andrew Charman. 2017. "Explicit Bayesian Analysis for Process Tracing." *Political Analysis* 25, no. 3 (July): 363–80.

———. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17, no. 1 (March): 154–67.

———. 2022. *Social Inquiry and Bayesian Inference: Rethinking Qualitative Research*. Cambridge: Cambridge University Press.

Goertz, Gary. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton: Princeton University Press.

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

Humphreys, Macartan, and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109, no. 4 (November): 653–73.

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

Mahoney, James. 2021. *The Logic of Social Science*. Princeton: Princeton University Press.

Qualitative and Multi-Method Research Fall 2022, Vol. 20.2 https://doi.org/10.5281/zenodo.8326443

Bayesian Reflections

Tasha Fairfield
London School of Economics

Andrew Charman
University of California, Berkeley

Social Inquiry and Bayesian Inference (Fairfield and Charman 2022) aims to share our enthusiasm for Bayesianism as a rigorous foundation for inference that can help strengthen and improve the natural intuition that qualitative scholars bring to their research. By way of introduction, Bayesian inference is a largely intuitive process that begins by assessing the *prior odds* on rival hypotheses—that is, how plausible we find one hypothesis relative to rivals—drawing on any relevant initial knowledge we possess. We proceed to gather evidence. We evaluate the inferential weight of the evidence by asking which hypothesis makes that evidence more expected, and how much more expected relative to rivals—the Bayesian term here is the *likelihood ratio* (sometimes called the Bayes factor). We then update to obtain *posterior odds* on our hypotheses—following Bayes' rule, we gain more confidence in whichever hypothesis makes the evidence more expected.[1]

We thank all the commenters for their thoughtful engagement with our ideas, many of which break with established approaches to inference in the social sciences. We are grateful for this opportunity to discuss, debate, and clarify various points.

Narrative Analysis and Bayesian Analysis

We concur with Bennett and Jacobs that there is ample scope for experimentation in how scholars incorporate Bayesian reasoning into qualitative research. As Bennett highlights, a central premise of our book is that many benefits can accrue from learning a bit about Bayesian inference, even if readers eschew the full machinery of Bayesian probability calculus. Yet we are also more optimistic about the role of explicit Bayesian analysis than we were at the outset of the project (Fairfield and Charman 2017), in part because we have a better sense of how often inferential errors can be made in case study analysis—in particular, taking evidence that is consistent with a theory to support that theory, without asking whether the evidence might be more expected under a rival. As we move toward more consciously structuring our thinking along Bayesian principles, it makes sense to write up and present that reasoning to readers, whether as a supplement to the case narrative that it informs, or potentially even as the centerpiece of a publication.

Jacobs is right to flag the disjuncture between traditional narratives and Bayesian inference, as well as the tradeoffs that scholars may face when deciding how to bring them together— these are indeed very different ways of presenting evidence and analysis. Yet research in the discipline commonly includes distinct components that do not necessarily fit neatly together—for instance, a multi-method design might include a formal model, a frequentist statistical analysis, and a case narrative.[2] As such, we would venture that presenting a narrative account alongside an overtly Bayesian analysis should not be seen as especially unusual or unwieldy. Moreover, we can begin to bridge the gap by recognizing the specific roles that narratives and Bayesian analysis play. Narratives allow authors to use their theory to explain their cases, while Bayesian analysis serves to explicitly test the theory by assessing how well it

1 For readers who are not familiar with Bayesian analysis, this video (https://www.youtube.com/ watch?v=Qvryz4RpTX0) may provide a useful introduction.

2 As we note in Chapter 9 (Fairfield and Charman 2022), these approaches juxtapose methods that draw on incompatible epistemological foundations.

outperforms salient rivals. To the extent that we value both endeavors—using an argument to explain empirics and testing the argument against rivals—including both components has merit. Word limits obviously pose constraints for journal articles, and here authors might well need to decide which component to emphasize in the main text. But scholars who wish to foreground a narrative may still be able to include illustrative Bayesian reasoning for a few key pieces of evidence in the main text while providing a more extensive Bayesian analysis as supplemental material. Alternatively, scholars might consider publishing a traditional narrative in one venue and a fully Bayesian analysis in another venue to reach different audiences.

As for Jacobs' point that narratives work best when all or most of the evidence supports the same hypothesis over rivals, whereas Bayesian analysis is ideally suited for handling less clearcut evidence, we agree. Explicit Bayesian analysis adds the most value when the evidence is nuanced and does not all weigh in favor of the same hypothesis (Fairfield and Charman, chap. 4, 164-66). And as Jacobs notes, these are also contexts in which a narrative account may be less useful or might convey more confidence in the leading explanation than the evidence merits. One of the main advantages of Bayesian analysis in fact is to keep us from overstating our confidence, or equivalently, to make us more aware of the uncertainty that surrounds our findings. Accordingly, we fully agree that in some situations it might make sense to prioritize explicit Bayesian analysis and abandon the narrative format. Our current project on covid origins adopts precisely that approach. Here we have a case for which the evidence is remarkably and notorious mixed—some observations weigh in favor of zoonosis, some favor a lab leak, and many observations that have been salient in public debate lend, in our analysis, little if any weight to either hypothesis. It is possible to write a seemingly coherent narrative from either a zoonosis perspective or a lab leak perspective, but even presenting both narratives in tandem, as if delivering opposing arguments to a jury,[3] does little to give readers a sense of which account is more plausible, and how much more plausible given what we know so far. A fully Bayesian approach that clearly delineates and analyzes each piece of evidence in turn is far better suited for systematically aggregating the inferential contribution of multiple evidentiary observations as well as avoiding confirmation bias (e.g., forgetting to ask whether evidence that ostensibly fits with one's preferred hypothesis might be as or even more expected under the rival hypothesis).[4]

At the same time, we would like to offer a few comments on the value of conducting and presenting a Bayesian analysis even when the evidence ostensibly lines up in favor of a leading hypothesis—considerations which lead us to hope that scholars will venture beyond the two Bayes-lite approaches that Bennett flags as most likely to take hold (simply harnessing knowledge of Bayesian probability to inform intuitive analysis of evidence, or evaluating likelihood ratios for just a few key pieces of evidence). First, it can be hard to discern how decisive the evidence actually is without focusing in on specific observations and asking how expected they would be under rival hypotheses. This point goes back to the above noted risks that case narratives may convey more confidence in our conclusions than the evidence warrants. Moreover, many narratives we have read do not do a good job of distinguishing argument and inference from empirics, and the evidence presented can be too vague or overly aggregated to evaluate its inferential weight. An explicit Bayesian analysis forces us to take greater care on these fronts and may in turn help us write better narratives. Second, readers may be more skeptical of the evidence for a claim than the author, so presenting a Bayesian justification for the weight that the author attributes to the evidence may help preclude disagreements, or at least provide a framework for discussing disagreements more productively. As emphasized in Chapter 7 (Fairfield and Charman 2022), we envision that one of the most important roles for explicit Bayesian analysis lies in structuring debates about inferences and making our analysis more amenable to scrutiny (here again we agree with Jacobs).

Process Tracing and Mechanisms

Process tracing and causal mechanisms have of course played a central role in the development of qualitative methods, and Bayesianism is often associated with process tracing in this literature. However, as Soifer highlights, our approach diverges from the notion that "tracing causal processes" or providing evidence for each step in a causal chain is adequate, or even necessary for inference to best explanation.[5] Setting out to "trace a causal process" can be an excellent way to inductively

3 See Chan and Ridley 2021, chap. 12.

4 In contexts that are not quite as ambiguous as the covid example, competing narratives, if carefully written, could prove useful for highlighting which observations fit well and which fit awkwardly with each theory, and for conveying where the respective stories seem more or less contrived or strained. But this approach would not be a substitute for systematic Bayesian analysis. Relatedly, we caution that an adversarial approach, which some have advocated, creates incentives for each side to overstate the strength of their conclusions, whereas the goal should be honest assessment of the uncertainty surrounding the conclusions.

5 We also emphasize that causal mechanisms are rarely directly observable; they are themselves a matter of inference.

devise theory.[6] But articulating a causal process inspired by the evidence we observe is not equivalent to testing our hypothesis.[7] Testing requires comparing a hypothesis to salient rivals and evaluating relative likelihoods of the evidence. This Bayesian perspective reveals that we should not limit the search for evidence to observations that bear directly on our theorized pathway from X to Y. Instead, we should recognize that any empirical observation which is more likely under one hypothesis relative to rival(s) contributes to updating, and we should seek out any evidence for which our hypotheses make divergent predictions.[8]

We would also like to offer some clarification regarding Soifer's musings on the role of mechanisms in our work that Bayesian inference is agnostic about the meaning of causation; it is compatible with whatever philosophy one wishes to adopt. Hypotheses could be formal models; they could invoke path dependence, complex conjunctural causation, or INUS causation; they could be either deterministic or probabilistic; they could be very specific about causal processes, or they could be less detailed, depending on the research agenda and the state of knowledge in the field. All we mean when we say that hypotheses should include a "causal mechanism" is that we should aim to clarify what kind of causal story we have in mind for how, why, and when some variables X_i lead to outcome Y.[9] That is, we should aim to give an explanation. We doubt that most scholars would disagree with that notion.[10] Even KKV (King, Keohane, and Verba 1994, 34) write that "explanation—connecting causes and effects—is the ultimate goal." Quantitative scholars may well tend to work with hypotheses that are more sparse on explanation or causal mechanisms, while qualitative scholars tend to offer more detail. And when working with nuanced and complex qualitative information from interviews, archives, or first-hand observation—which is the central concern of our book—we do indeed need to articulate hypotheses with enough specificity to be able to "mentally inhabit" the corresponding world and reason about what observations would be more expected or less expected. As such we agree with Soifer's interpretation that expounding causal processes or mechanisms serves to make our hypotheses more precise. It is worth emphasizing that specifying hypotheses can be an iterative process; we may start a research project with rather bare-boned hypotheses and revise them to provide more detailed explanations or causal pathways as we learn more. Our Bayesian approach is accordingly compatible with research that begins by "soaking and poking," with only tentative initial ideas about possible explanations.

Cross-Case Analysis

Although Bayesianism has most often been associated with process tracing and within-case analysis, we argue that in a Bayesian framework, there are no fundamental distinctions between within-case analysis and cross-case analysis. Whether we are studying a single case or multiple cases, all evidence contributes to inference in the same manner—by evaluating likelihoods under rival hypotheses. To recapitulate our approach, a well-articulated hypothesis should include a statement of its scope, beyond which it makes no predictions. Observations from any case that falls within the stated scope of the hypotheses under comparison then contribute some weight of evidence to the inference. In the same way that inferential weight accumulates for each evidentiary observation pertaining to a single case, the inferential weights of multiple pieces of evidence aggregate across cases and contribute additively to the posterior log-odds on the hypotheses.[11] Inferences are always provisional and comparative, in the sense that (i) posterior odds reflecting what we have learned from cases already examined become "prior odds" when moving forward to consider new cases, but what we discover in new cases may well change our view about the relative plausibility of alternative explanations, and (ii) we are always free to devise new or refined hypotheses to compare.

Four points may help to clarify our approach with respect to ~~regarding~~ Bennett's and Jacobs' queries about learning across cases. First, hypotheses, including their scope conditions, must be propositions with well-defined, if imperfectly known, truth values. A scope

6 Tracing a causal process may also be an effective way to deploy theory to explain a case.

7 In our view, some of the literature on process tracing and causal mechanisms conflates hypothesis generation with hypothesis testing (see *Qualitative & Multi-Method Research* 18(2)), while qualitative research that invokes process tracing as its methodological foundation often engages less in theory testing than in proposing a theory and using it to explain a case.

8 Nor do we necessarily need to examine evidence pertaining to every granular step in the causal chain, particularly if the hypotheses under consideration do not make strongly divergent predictions at some steps.

9 There is of course an ample literature that debates what exactly causal mechanisms are and what their relation is to inference (e.g., *Qualitative & Multi-Method Research* 14(1)), which we regard as largely beside the point from a Bayesian perspective.

10 Some might however take issue with our use of the term "causal mechanism," which is sometimes associated with deterministic causation, whereas we expect that probabilistic models of causation are more realistic and useful for most social science contexts.

11 Here we are invoking the log-odds version of Bayes' rule: the posterior log-odds equal the prior log-odds plus the net weight of evidence (Fairfield and Charman 2002, chap. 4).

condition itself is a logical proposition with some binary truth value that defines the contexts in which the hypothesis makes predictions, versus contexts in which the hypothesis makes no predictions at all. We may have epistemic uncertainty as to whether a case satisfies the stated scope, in that we do not have enough information about the case to know for sure.[12] But we do not take scope conditions to involve any intrinsic, aleatoric uncertainty.[13] Any uncertainty about "the degree to which conclusions travel across the domain of theoretical interest" (Jacobs, this symposium) is reflected in the probabilities of the articulated hypotheses with their stated scope conditions—which are part and parcel of the hypotheses themselves—just as these probabilities reflect uncertainty about any other aspects of the hypotheses, namely, the causal logics or mechanisms they propose.

Second, while hypotheses do need to contain clearly articulated scope conditions before applying the Bayesian inferential apparatus, scope does not need to be rigidly determined at the outset of research. As we learn more, we can always revise the scope conditions in our hypotheses to either pose them at higher levels of generality or restrict their predictions to narrower contexts, in accord with Bennett's observation that our understanding of scope can change substantially over the course of research. We view the complications Bennett emphasizes on this front as part of the usual give-and-take of iterative theory building and testing. Analyses can be revisited, observations can be analyzed differently or more deeply, different parts of our background information may become more or less relevant, and both theorized scope conditions and causal mechanisms can be tweaked.[14]

Third, priors and posteriors are associated with the hypotheses under comparison and necessarily match the stated scope conditions that the hypotheses articulate. Regarding Bennett's query about case-specific priors and Jacobs' question about distinguishing posterior beliefs about cases for which we have observed evidence from posteriors about cases from which we have not yet observed evidence, our response is that case-specific hypotheses have case-specific priors and case-specific posteriors; whereas hypotheses with broader scope have priors that are informed by all salient background knowledge possessed about each of the cases within its scope, and posteriors that draw on all evidence learned from any case within the scope. Medical examples like the one Bennett introduces are best understood as using (rather than testing) theories to diagnose or make prognostic predictions for an individual case (a patient),[15] as are examples of generating and assessing hypotheses about an individual case (e.g., the patient has ovarian cancer vs. irritable bowel syndrome). As for our social science example on democratic mobilization, when comparing hypotheses that are scoped to make predictions throughout Southeast Asia, logically speaking we cannot ask about priors or posteriors that apply only to some subset of Southeast Asian countries vs. priors or posteriors that apply to some other subset thereof. That is, different cases that fall within the scope of the hypotheses under comparison cannot have different priors or posteriors.[16]

Fourth, a hypothesis that makes predictions within a given theoretical domain or scope need not assert causal homogeneity across the entirely of that domain. A hypothesis can apply one causal logic within some subregion of its scope space and another distinct causal logic within some other subregion of its scope. "Patchwork" hypotheses of this sort assert causal heterogeneity, while still making predictions across all cases within their scope. (While some readers may tend to associate scope with a particular causal mechanism or casual logic, we emphasize again that the scope of a *hypothesis* is simply a statement about the contexts in which it makes predictions of any kind, vs. those in which it makes no predictions.) As we study more cases or expand the scope of our hypotheses, we may well want to consider causally heterogeneous patchwork hypotheses, as per Bennett's expectation (this symposium) that "in social life there are few simple hypotheses with broad

12 For example, values of some socioeconomic indices may not have been measured or reported with sufficient precision to determine whether a country has crossed specified thresholds.

13 Scope conditions involving categories like "developed countries" or "social democracies" are not probabilistically uncertain but rather semantically fuzzy, until the associated concepts are defined more precisely.

14 Fairfield and Charman 2022, chap. 5 provides guidance on iteratively adjusting scope conditions; see chap. 10 on iterative research.

15 We might imagine inputting the patient-reported symptoms, case history, and results of physical examination and diagnostic tests into some sort of logistic regression model, or neural network, etc., in order to generate a posterior predictive probability distribution over possible diagnoses. Any reasonable model should of course make use of suitable priors or "base-rates" appropriate to the medically relevant reference classes to which the individual belongs, as well as more case-specific information as it becomes available.

16 See also Fairfield and Charman 2022, appendix 12.D.

scope."[17]

We can now address one of Jacobs' central concerns about how the distribution of evidence across cases matters for updating. Referencing the democratic mobilization example, he worries that when aggregating weights of evidence, our approach allows "no distinction to be made between observing ... three highly probative ... pieces of evidence in favor of the communal elites hypothesis (relative to its rivals) within a single case, on the one hand, and observing ... three highly probative pieces spread across three separate cases, on the other hand" (this symposium). If we are comparing two causally uniform hypotheses, H_{CE} = *Slater's communal elites causal logic operates throughout Southeast Asia* vs. H_{ED} = *Economic decline sparks democratic mobilization throughout Southeast Asia*,[18] then indeed it does not matter whether the evidence comes from one case or is spread across three cases, because under either theory, the mechanism is asserted to be the same across all Southeast Asian cases. However, suppose that we compare H_{CE} to a more complex, causally heterogeneous hypothesis $H_{CE/ED}$ = *The communal-elites causal logic operates in the Philippines and Vietnam, whereas economic decline instead sparks democratic mobilization elsewhere in Southeast Asia*. Finding three pieces of evidence from the Philippines that strongly support the communal elites causal logic over the economic decline causal logic would then fail to discriminate between H_{CE} and $H_{CE/ED}$, whereas if one piece of evidence of similar strength for the communal-elites causal logic were found in each of three cases—the Philippines, Vietnam, and Burma, that evidence would support H_{CE} over $H_{CE/ED}$—albeit with only the evidence from Burma contributing inferential weight. If instead the third piece of evidence from Burma favored the economic decline causal logic over the communal elites logic, then the three pieces of evidence taken together support $H_{CE/ED}$ over H_{CE}—we might then say that based on our knowledge so far, the communal elites *causal logic* does not generalize beyond the Philippines and Vietnam, but $H_{CE/ED}$ nevertheless provides a viable (if tentative) explanation for democratic mobilization in all Southeast Asian countries.

Accordingly, it is important to remember that whether and how evidence in one case is informative about other cases depends on the hypotheses under consideration. If our inference from the Philippines involves hypotheses scoped to include only this country, then however strongly the posterior odds favor one or the other explanation, those hypotheses would make no predictions whatsoever about what we ought to find in Burma. If we then *tentatively* expand the scope conditions to include all of Southeast Asia, the hypotheses now do make predictions about how things should work in Burma, and *if* these are the only hypotheses under consideration, then evidence collected from the Philippines will indeed shape our current views about the leading hypothesis for understanding democratic mobilization not only in that country, but also in Burma. But if we include hypotheses that postulate operation of different mechanisms in the Philippines and Burma, then it will become important to also look at evidence from Burma.

Generalization then does not happen automatically or by fiat in our approach—we do not get something for nothing, as Jacobs fears. Instead, generalization involves hypothesizing and testing—we *compare* rival hypotheses that make predictions for some shared set of cases. Any background knowledge we have about homogeneity of cases should inform how we craft hypotheses and evaluate their prior odds; updating will depend on what evidence materializes and how and where the predictions of our rival hypotheses diverge.[19]

Some of the skepticism that Jacobs and others have expressed about our approach to generalization may stem from not fully appreciating the conditional and contingent nature of our Bayesian reasoning.[20] We cannot emphasize enough that both our theories and the credences we hold in them are provisional. We are always free to revise hypotheses, whether by changing the causal logic or altering the scope. As we consider a broader set

17 A caveat related to Occam's razor (Fairfield and Charman 2022, chap. 6) applies here. By any sensible measure of complexity, there will be exponentially more complex theories than simple ones that might in principle be considered. This has important consequences. Even if we put more prior probability on the class of complex hypotheses than the class of simple ones, any one complex hypothesis would tend to have lower prior probability than any one simple hypothesis, because there are so many more possibilities of the former class compared to the latter. Accordingly, our best strategy is to start by considering simpler theories, only adding complications or elaborations as necessary, as the simpler theories falter. And by reflecting on how simpler theories fail, we often find hints about how to improve them.

18 See Slater (2009) and Fairfield and Charman 2022, chap. 5.

19 Both Jacobs and Bennett appear to want to presume causal heterogeneity unless there is positive evidence otherwise. But Occam's razor suggests the opposite strategy. As for "building the researcher's beliefs about heterogeneity directly into the likelihoods of the evidence" (Jacobs, this symposium), we contend that this is a job for theory. A carefully articulated and scoped hypothesis builds conjectures about the homogeneity or heterogeneity of cases into the likelihoods of possible evidence.

20 We suspect that some of Jacobs' concerns may also reflect a commitment to working with a potential-outcomes framework and assigning cases to latent causal types (Humphreys & Jacobs 2015), whereas our approach focusses directly on causal explanations articulated in rival hypotheses (see Fairfield and Charman 2022, chap. 9, 395-96). We look forward to further discussing the distinctions between our approaches in a future setting.

of cases, beyond proposing patchwork hypotheses, we might devise new hypotheses that endogenize what we previously considered to be a binary scope condition, so that it becomes part of the (possibly probabilistic) causal logic itself, perhaps as a (no longer binary) moderating variable (Fairfield and Charman 2022, chap. 5, 204-17). After each iteration of hypothesis refinement, we apply the Bayesian apparatus to ask which hypothesis among a set of comparably scoped alternatives provides the best explanation in light of the evidence in hand so far. As data accumulate, a given explanation may gain or lose plausibility in relation to rivals that might posit different or more heterogeneous explanations.

A Unifying Logic for Inference

We are happy to embrace Soifer's characterization of our book as a Bayesian version of KKV (King, Keohane, and Verba 1994)— we share a similar overarching goal of providing a unified approach to inference that applies to both qualitative and quantitative data (Fairfield and Charman 2022, chap. 9), and "observable implications of theories" do indeed play a central role in our framework. We would characterize KKV as a frequentist-inspired perspective on qualitative research, which we find problematic because according to its own foundational principles, frequentism can only be used to analyze stochastic data. Bayesianism is the only natural and logically rigorous inferential framework that can accommodate both qualitative and quantitative evidence—regardless of what type of information is in hand, inference proceeds according to the same underlying principle: evaluate relative likelihoods for the evidence under rival hypotheses. As such, what Soifer characterizes as a purely practical or "omnivorous" approach for using any informative data to test theories actually rests on deep foundational principles (Fairfield and Charman 2022, chap. 2). Furthermore, Bayesianism is ideally suited for addressing KKV's (King, Keohane, and Verba 1994, 32) central critique of qualitative political science: "the pervasive failure to provide reasonable estimates of the uncertainty of the investigator's inferences." Bayesian probability is an extension of Boolean logic to contexts of uncertainty and limited information; inferences are expressed as posterior odds that characterize how much confidence we hold in a hypothesis relative to rivals given the evidence in hand, or equivalently, how much uncertainty remains regarding which hypothesis provides the best explanation.[21] A Bayesian framework also clarifies that what matters is not how many empirical observations line up with our theory, but rather the *relative likelihood* of the evidence under rival theories.[22]

While we are indeed pushing back on the now conventional QMMR understanding that qualitative and quantitative research follow different logics of inference, we agree with Goertz and Mahoney that these research communities have been characterized by different cultures of inference. But when comparing conventional quantitative research to in-depth qualitative research, we would argue that the cultural difference is marked by frequentism versus intuitive Bayesianism. We suspect that this epistemological mismatch (even if not explicitly recognized as such at the time) is what motivated much of the reaction from qualitative scholars against KKV's prescriptions, some of which impose impractical constraints that are not necessary within a Bayesian framework, including the stricture of testing theory with new data that was not used to inspire or refine the theory.[23] Bayesianism by contrast gives a solid mathematical foundation for iterating between data collection and theory refinement, as well as many other common practices in qualitative research that are not justifiable within a frequentist framework.[24] We hope that qualitative scholars will find these foundations empowering. Bayesian updating in our experience mirrors the way many scholars naturally approach research. And putting our approach into practice involves very little math. Even for those who choose to use the quantified version of Bayes' rule with log-odds, nothing more than addition and subtraction is required.

As to Soifer's perception of a tension between our commitment to qualitative research and our premise that Bayesianism provides a unified inferential framework, we instead see these matters as closely related and complementary. Recognizing Bayesian probability as a universally applicable framework places qualitative evidence on much more equal ground relative to quantitative data and experimental evidence and should thereby help to clarify and revalue the contribution of qualitative information to causal inference, which we understand as inference to best explanation. As discussed in Section 2, Bayesianism imposes no constraints on the notion of causation that hypotheses embrace, so we do not anticipate the particular discord that Soifer contemplates, although we of course recognize that

21 Interestingly, KKV (King, Keohane, and Verba 1994, 32) end their section on "Reporting Uncertainty" by encouraging one to ask: "How much ... would you wager" or "At what odds," which is inherently Bayesian.

22 An additional distinction is that in contrast to frequentist requisites, in a Bayesian framework, all observable implications need not be listed in advance of data collection (Fairfield and Charman 2022, chap. 10).

23 See for example Ragin (1997, 3).

24 See especially Fairfield and Charman 2022, chaps. 10, 12.

adopting a Bayesian framework would require many quantitative scholars as well as many qualitative scholars to change their research practices.

Bouchat expresses more substantial doubts about the value of Bayesianism for equalizing the role of qualitative and quantitative research, based on a claim that quantitative critiques of qualitative research now rest not on issues of inference, but rather on issues of data selection, case selection, generalizability, informativeness of data, and "data validity" or "data quality." First, we fail to see how "data validity" in the sense of measurement validity could underpin quantitative critiques of qualitative research, since if anything qualitative scholars would seem to have the advantage on this front, thanks to in-depth case knowledge.[25] Second, and most importantly, we would counter that none of the other considerations can be separated from inference. Scholars critique these aspects of research because they matter for inference, but how and to what extent they matter depends on the espoused methodology. Frequentism and Bayesianism treat case selection and other aspects of research design very differently. They handle bias differently. They understand and conduct generalization differently. And they evaluate informativeness of data differently. Judgements about "data quality" are likewise directly linked with the ability to draw reliable inferences, so this too is ultimately a methodological question. To focus on any of the particular concerns Bouchat mentions while overlooking methodological distinctions between frequentism (the dominant framework for quantitative political science) and Bayesianism, or any other inferential approach that qualitative scholars have espoused, is to miss the underlying source of disagreements and tension. Clarifying these distinctions is a central aim of our book.

Bouchat (this symposium) goes on to say that the desired goal should be to "treat evidence derived qualitatively as equal with that measured and collected quantitatively"—but we contend that Bayesianism does just that, precisely by virtue of drawing inferences from all data in same manner. Of course, not all evidence will be equal in terms of its inferential import, but inferential import depends on how strongly the evidence in hand discriminates between rival hypotheses, not whether it is qualitative or quantitative. As for the assertion that Bayesianism "does not at all resolve ... what qualifies as good data," we are perplexed, considering that Bayesianism to our minds provides a clear and straightforward answer: "good" data are *informative* data, namely, any observations that are more expected under one hypothesis compared to rivals. The more divergent the likelihood of the data under rival hypotheses, the more informative the data, and hence the "better" the data. If the data are noisy or imperfect, then Bayesians can and should take that into account. If there is a question about the validity of a measurement (i.e., whether it captures the concept or variable of interest), Bayesians can and should take that into account as well, by conditioning on the raw data as they are, not on the value of some variable that we hoped to measure but did not. So again, we do not see concerns about data validity or data quality as either a fundamental source of difference between quantitative and qualitative research, or as a challenge to our argument that Bayesianism serves as a universal framework for inference that revalues the contribution of qualitative evidence.

A second leg of Bouchat's critique, in our perhaps imperfect understanding, is that by espousing an objective Bayesian framework rather than fully embracing subjectivity, we necessarily undermine qualitative research, which inherently involves subjectivity. We find this reading counter to the intent and substance of our book; we of course fully agree that qualitative research "does not lack in its scientific value by leveraging data or insights that defy easy quantification" (this symposium). Throughout, we acknowledge that subjective inferences are necessary in practice, and we emphasize that quantitative social science is no exception—not only to the extent that it draws on qualitative information that has been imperfectly quantified to construct datasets, but also through the many decisions made when elaborating models that necessarily require scholarly judgement.[26] But our goal is articulate principles and illustrate practices that can help social scientists to reason as rationally and objectively as possible about the way the world works. Understanding and following Bayesian principles helps our subjective judgements better approximate the ideal of rational inference, while simultaneously allowing us to leverage all the information in nuanced, detailed, qualitative evidence. We do not think that working to minimize subjectivity in inference cedes ground to any claims that quantitative research is superior due to greater objectivity—again, we explicitly argue against the notion that objectivity vs. subjectivity distinguishes quantitative from qualitative social science (Fairfield and Charman 2022, chap. 9, 440-45)—or that it undermines our vision for Bayesanism as a unified inferential framework. We

25 Bouchat (this symposium) further mentions validity "in the substantive sense" of "studies that do not identify causation"—we are not sure exactly what is meant here, whether it be all qualitative research, which by frequentist quantitative standards cannot produce causal identification, or specifically qualitative research that does not focus on explanation. But it is worth emphasizing that our book primarily speaks to qualitative research that aims to make causal claims.

26 We return to this point in Section 5.

return in Section 6 to clarify some specific points about priors and knowledge accumulation that might have fostered perceptions to the contrary.

Formalization and Analytical Explicitness

While our approach to inference is guided by the formal apparatus of probability theory, we do not, as Jacobs notes, formalize the derivation of likelihood ratios. Formalization would require devising a statistical model (e.g., regression-like structural equations, input-output tables, or an instantiation of a DAG) capable of producing precise numerical likelihoods in an algorithmic way for every possible piece of evidence that might be observed during data collection. Instead, we quantify relative likelihoods only for the empirical observations that do turn up, once they are in hand, based on informal but careful verbal reasoning about the predictions that our plain-language hypotheses suggest.[27]

We take Jacobs' point that formalization and objectivity are conceptually distinct, and we recognize that describing formalization as creating a "veneer of objectivity" may not have adequately conveyed why we prefer to reason informally about likelihood ratios. To clarify, we contend that formalization of the sort Jacobs has in mind is essentially impossible when working with the kind of detailed qualitative evidence that is the central concern of our book—that is, we do see "fundamental limits" to formalization in this context.

The problem lies in that formalization requires specifying probabilities for all possible empirical observations in advance, but we cannot hope to even envision all such possibilities when the evidence in question involves open-ended responses from expert informants, passages from archival sources, accounts from newspapers, firsthand observations of human behavior, visual information from campaign ads, and so forth. To illustrate the scale of the problem, consider information that a scholar might elicit from an expert informant during an interview. If the informant gives a three-sentence reply to just a single question, there may be on the order of 10^{200} possible responses (taking into account the average length of a sentence in English and ignoring non-verbal cues) that would need to be enumerated and then assigned likelihoods.

Any effort to formalize hypothesis testing with this kind of qualitative evidence would require massive coarse-graining of potential observations into a manageable number of categories. But details in the evidence (e.g., tone of voice, body language, identity of the informant, context in which remarks are made) can matter greatly for likelihoods, so the probability that the model assigns to any one of the coarse-grained evidentiary types it specifies may not be an adequate approximation for the likelihood of any concrete qualitative empirical observation that turns up. The coarse-graining required for formalization in essence throws away relevant information in the evidence and distorts the conclusions. In contrast, our approach avoids what we see as the unnecessary and near impossible effort of assigning probabilities to the myriad possible evidentiary observations that might have materialized but did not, while allowing us to use all the information in our evidence.[28]

Subjectivity inevitably enters our informal approach when reasoning about which of one or more rival hypotheses (expressed in ordinary language) makes an evidentiary observation more expected, and in assigning numerical values to represent our judgements about evidentiary weight.[29] In logical Bayesianism, objective probabilities are determined exclusively by the information available. Subjective probabilities draw not only on the information available, but also on judgement, which should be informed by expertise and experience, but will also involve some degree of arbitrariness. While the guidance in our book aims to help subjective probabilities better approximate the logical Bayeisan ideal, the fact remains that there is no strictly objective way to quantify probabilities for complex, nuanced, inherently qualitative information about the socio-political world.

But we maintain that our approach makes this subjectivity transparent and invites discussion among scholars who may think differently, which in turn facilitates consensus building, or at least clarification of where any why scholars disagree. As such, we would say that we achieve the same goals without a formal model that Jacobs highlights in writing that "a model representing the researcher's beliefs about how the world works, and from which the likelihoods are then derived, makes explicit elements of the analysis that will otherwise remain implicit." In our approach, assigning some qualitative observation a weight of 10 dB in favor of H_1 vs. H_2 clearly conveys our degrees of belief to readers. And we accompany this quantitative judgement up front with a written explanation for why we consider the evidence to be moderately more expected under H_1 vs. H_2. In contrast, the formalized approach advocated by Jacobs to our minds hides the researcher's views

27 We are referring here to our approach for log-odds updating (Fairfield and Charman 2022, chap. 4), which we call "explicit Bayesian analysis" (for lack of a better term), as opposed to "heuristic Bayesian reasoning" (Fairfield and Charman 2022, chap. 3), which applies the same thought process but stops short of quantification.

28 See Fairfield and Charman 2022, chap. 10, 479, 495-98.

29 Subjectivity of course also enters when evaluating priors (see Section 6).

within the intricacies of the model, in a way that makes it more difficult for readers to understand, evaluate, and critique—at least when the evidence involves inherently qualitative information. When applying our approach, if another scholar asks why we deemed the weight of evidence to be 10 dB rather than 15 dB, we can have a conversation on the spot, which may lead us to better articulate our reasoning, specify our hypotheses more clearly, or revise our views. When employing formal models of the sort proposed by Humphreys and Jacobs (2023), an analogous discussion would involve questions about distributions over latent variables or parameters, the precise form of structural equations, or other highly technical attributes of the model that are more difficult to connect to the substantive meaning of a theory and the evidence in hand.

We of course do not object to formalization in all contexts. But for in-depth qualitative research, formalization would involve replacing a manageable number of subjective but direct judgements about likelihood ratios for observed evidence with a vast number of ultimately subjective choices about technical intricacies of the model.[30] We envision few benefits in terms of explicitness or transparency to embedding probabilistic judgements in multiple layers of parameterizations with limited interpretability when the inferences we care about are the relative plausibilities of rival theories that provide distinct explanations for socio-political phenomena. This is what we had in mind when writing that formalization "simply pushes the subjectivity back deeper into the model" (Fairfield and Charman 2022, 442).

Lastly, we do not agree with Jacobs' remark that "formally deriving priors and likelihoods from a single underlying model forces internal consistency among the inputs to Bayesian analysis," (this symposium) for the simple reason that a model itself cannot provide all of its own priors. For DAGs of the sort discussed in Humphreys and Jacobs (2023), each node in the graphical model will typically require many *exogenous* inputs determining prior probabilities over various nodal types. More generally, hierarchical modeling can push the exogenous probability inputs into deeper layers, but that does not circumvent the need to make largely arbitrary choices about prior distributions for hyper-parameters which may influence the observable predictions of the model in ways that are difficult to discern. And what we consider the most important prior probabilities for theory testing, namely those specifying relative plausibilities for the overarching model families that offer competing explanatory frameworks (e.g., specifying distinct DAG topologies), can never be regarded as part of the model—hypotheses cannot assert their own degree of plausibility.

We believe our differences of perspective on these points stem from the distinct research contexts we focus on as well as our orientation toward hypothesis testing. As already discussed, our work focusses on analyzing open-ended, detailed qualitative observations, whereas Humphreys and Jacobs (2023) apply their formalized approach primarily to moderate numbers of variables that assume only a moderate number of values. Furthermore, we engage in theory testing by comparing rival hypotheses, which would be the heuristic or qualitative analog of comparing distinct model families, whereas Humphreys and Jacobs largely focus on what would be considered parameter estimation in standard statistical parlance, along with other inferences within a single chosen model family.[31]

Priors and Knowledge Accumulation

While the role of prior probabilities in Bayesian inference does constitute a major departure from frequentist frameworks, the importance of priors is sometimes exaggerated by critics. In our view, Bouchat's claim (this symposium) that the guidance our book offers for Bayesian reasoning "only makes sense conditional on the establishment of prior probabilities" is similarly exaggerated. We instead hold that weights of evidence merit much greater attention than priors in qualitative research that aims to bring new evidence to light.[32] Moreover, the same guidance provided in our book would directly apply to research agendas that aim to systematically construct informed priors or characterize the existing state of knowledge in a field. Before elaborating these points, we briefly review our approach to priors.

Recall that probabilities within objective, or logical, Bayesianism are degrees belief determined by states of

30 In Humphreys and Jacobs' (2023) potential-outcomes framework, the total number of causal types, and the parameterizations associated therewith, grow super-exponentially with the number of distinct values or categories that the independent and dependent variables can assume. While the growth in complexity can be partly tamed by a choice of a particular DAG topology, full formalization will unavoidably require an enormous number of largely subjective decisions to give concrete shape to the probability model.

31 Translated into Humpheys and Jacobs' (2023) framework, what we are doing would involve comparing distinct DAG topologies involving substantially different nodes or different connections between nodes.

32 Priors matter more for quantitative research involving parameterized models. Here we are interested in prior odds on competing theories (or model families).

knowledge.[33] Accordingly, the aspirational goal would be to incorporate all relevant initial information (and nothing else) into our prior odds. In principle, we would go back to a state of minimal knowledge or ignorance that justifies equal odds on hypotheses of comparable complexity, and then build up to our present state of knowledge by employing Bayes' rule, effectively incorporating all of our initial information as evidence (Fairfield and Charman 2022, chap. 3, 96). Practically speaking, however, we usually have too much initial knowledge to carry out this procedure, short of turning the construction of priors into the sole focus of research. For work that aims to bring new evidence to light, we will have to make do with subjective approximations to the logical Bayesian ideal, in that we will need to use judgement to guide us rather than attempting a full and systematic accounting of background information. As such, we suggest two options: (1) articulate informed priors as best as possible, explaining how key elements of background knowledge motivate these judgments, or (2) just start from equal odds on the salient hypotheses, which focusses attention on the evidence at hand and in essence allows readers to supply their own priors. Whether starting with informed priors or indifference priors, it is sensible and straightforward to conduct sensitivity analysis by exploring the import of different priors, including priors that anticipate the reaction of skeptical readers whose background knowledge might lead them to prefer a rival hypothesis over the author's favored argument. Such sensitivity analysis is almost trivial when working with the log-odds form of Bayes' rule.

We now turn to clarifying several points with regard to Bouchat's critique. First, indifference priors in qualitative Bayesian reasoning are not meant to "reflect impartiality and objectivity." As explicated above, true objectivity would involve systematically incorporating every relevant element of the scholar's background knowledge into their prior odds, which as a practical matter may be impossible in most social science contexts because we simply possess too much background knowledge. Instead, using indifference priors in contexts where we do not actually find ourselves in an initial state of ignorance is a pragmatic recommendation to address the reality that readers will inevitably bring their own very different priors, based on very different background information, to bare on our work. Given this reality—and stressing how dramatically background knowledge and hence prior beliefs can vary among scholars—we contend that the most important task is to focus on the inferential weight of the evidence we are contributing to the literature. The greater the weight of evidence in hand, the less priors will matter for posterior judgements, and scholars who start with different priors may still end up favoring the same hypothesis in light of the evidence. And even if priors remain poorly specified or contested, carefully analyzing the weight of the evidence in hand can still make an important contribution to knowledge accumulation. Furthermore, by reporting weights of evidence, or equivalently, posterior log-odds based on indifference priors, authors and readers can immediately discern what strength of prior belief would be needed to overcome the import of the new evidence.[34]

Second, we have no objections to subjective priors as a heuristic, so long as they aim to reflect the scholar's empirical background knowledge, rather than desires about how the world ought to work or empirically unjustified preferences for a pet theory—these kinds of considerations are subjective in a non-scientific sense, as opposed to subjective in the sense of varying across individuals who simply possess different information.[35] Values, desires, and personal preferences can certainly guide the choice of research questions and ethical research practices, but they should not affect inferences from empirical evidence.[36]

As for the alternative of eliciting priors and pooling opinions from experts, we recognize that this is an active area of scholarship within subjective Bayesianism, and

33 Strictly speaking, probabilities cannot be "measured," since they are epistemological rather than empirical. Bouchat's comments, however, frequently refer to measurement, in ways that leave us unsure of the intended meaning. For example, Bouchat reads us as advocating that scholars "conform as much as possible to objective aims, measurement, and likelihood specification." This characterization is correct on the first and last accounts, but measurement in this context does not comport. Our book is not about measurement; it is primarily about inference with qualitative evidence, which involves assessing likelihood ratios rather than measuring or scoring some variable or concept.

34 Notice that using indifference priors is mathematically identical to dropping priors and simply reporting weights of evidence, which is a common Bayesian convention.

35 The pages Bouchat highlights as evidencing our outsized "concern about undue influence, subjectivity, and bias" in priors involve our response to particular issues that *other scholars* have posed to us, specifically regarding our treatment of iterative research. We view their concerns as legitimate but easily addressed (Fairfield and Charman 2022, chap. 10, 491-92).

36 Regarding Bouchat's interest in how to rigorously establish informed priors, beyond our guidance to provide an explanation for one's view that highlights the most salient elements of one's background knowledge, our advice would be to identify and analyze concrete empirical information from the literature reviewed, and then build up from ignorance priors to informed priors using Bayes' rule. Short of conducting a Bayesian meta-analysis (see below), which would be a research project unto itself, there is no "off the shelf" instruction manual for how to circumscribe this process to make it a feasible task—scholars would have to exercise judgement and explain their decisions, just as they do when arriving at informed priors as per our more informal guidelines.

we acknowledge that this kind of approach may be useful for some research agendas. However, there is no widely accepted algorithm for these tasks that can be fully justified with objective Bayesian principles. We also caution that even a rigorous methodology for aggregating potentially divergent expert opinions may run up against the limitation of experts who are not themselves Bayesian reasoners. While expert opinions draw on expert knowledge, experts may not arrive at their opinions via any sort of coherent Bayesian principles. And it is far from clear whether imperfections in individual scholars' reasoning can be averaged away through the aggregation process, especially if "conventional wisdom" leads to positively correlated errors.

From a logical or objective Bayesian perspective, we would ideally want to pool experts' *empirical knowledge*, rather than experts' opinions, and then carefully analyze that knowledge to arrive at relative odds on salient hypotheses. At least in principle, this could be done by training experts in Bayesian reasoning and holding workshops where knowledge is shared, analyzed, and debated (along the lines of the research agenda Bennett mentions). Importantly, notice that this process would involve treating what would otherwise be background information as evidence, and would thus become identical to assessing and scrutinizing weights of evidence as per the guidelines in our book, with a focus on known facts within a research community rather than new evidence obtained through original research.[37]

Turning to knowledge accumulation, Bayesianism is an ideal framework for learning both across different components of a single study and across distinct studies. Whatever the data source or type, weights of evidence accumulate additively,[38] and prior log-odds add to the total weight of evidence to yield posterior log-odds. In the first context, scholars conducting, for example, Bayesian analysis of a quantitative dataset followed by case studies (or vice versa), can employ their posteriors from the first component of research as their priors for the second component of research. In this manner, knowledge accumulates naturally across aspects of the research that draw on distinct kinds of data, without recourse to different methodologies that draw on incompatible epistemologies and produce findings that are not easily integrated. Here we are not sure what to make of Bouchat's suggestion that our recommendation for scholars to use their own background knowledge and priors undermines our vision for Bayesianism as a unified inferential framework, considering that learning across components of a study proceeds in the manner described above regardless of how priors for the first component of research were generated. If the priors and weights of evidence are reported separately, then readers can substitute their own priors, or if they wish, try to formulate some sort of consensus prior for the relevant research community.

Regarding knowledge accumulation more broadly—not just across different components of a research project, but across distinct studies, perhaps aiming to draw on all relevant published literature—one enters the realm of what we might call meta-analysis. While this is not the focus of our book, the same principles and guidance apply at this level. Bayesian macroknowledge building or meta-analysis would be straightforward if all studies in the literature reported Bayesian weights of evidence with respect to leading rival explanations: weights of evidence would then be additive across studies in the same way that they are additive within studies.[39] But reporting weights of evidence has not been standard practice in social science.[40]

Given this status quo, a careful meta-analysis designed to assess the state of knowledge in a field would require (1) devising a common set of explanatory hypotheses to compare that includes the leading arguments under debate, (2) extracting concrete empirical evidence from literature in the field, and (3) conducting Bayesian inference. While we see ample potential here for major contributions to social science, this kind of project would involve a very substantial amount of effort. For qualitative research on, say, state building, one would need to employ a team of trained scholars, and ideally engage experts in a process of scrutiny, adjudication, and consensus building.[41] If a project of this sort proved achievable, scholars could then employ the resulting posteriors as priors for additional research on the topic. But significant challenges remain, in that once someone invents a new hypothesis to test, rigorously speaking, they would have to go back through the entire body of evidence considered in the meta-analysis to construct prior log-odds for the new hypotheses relative to rivals.[42]

37 Alternatively, one could interview or poll domain experts about competing theories and try to use these responses as *testimonial* evidence, but the likelihoods would be extremely challenging to assess.

38 Provided that allowance is made for possible logical dependency in the data given the hypotheses.

39 Again, modulo any dependency considerations. And additional analysis, revisiting the original data, would of course have to be conducted for new hypotheses that were not previously assessed.

40 An obvious first step is to train scholars in Bayesian reasoning, which is the purpose of our book.

41 We have been actively looking into opportunities for conducting precisely this kind of research and would be happy to hear from any interested potential collaborators.

42 Fairfield and Charman 2022, chap. 10.

If a scholar's primary goal is to contribute new evidence to the debate, then we would reiterate the advice in our book: rather than undertaking the mountain of effort needed to systematically incorporate all relevant background knowledge from existing literature, articulate priors that aim to reflect the most consequential elements of your own background knowledge, and then focus on evaluating new evidence.

As for accumulating knowledge across disciplines, the same principles expounded in our book are directly applicable here too—Bayesianism is a natural framework for knowledge accumulation in all contexts. Our current research on covid origins demonstrates how an informal Bayesian framework can be used to organize and analyze diverse kinds of evidence produced by multiple fields of inquiry, ranging from genomic information and epidemiological evidence to information from observational field work, testimonial accounts, and journalistic reports. Our research has involved reviewing literature and interviewing expert informants across disciplines as diverse as virology, genomics, zoology, medicine, geography, and political science. The same caveats expounded in our book apply in this context as well. Quantifying weights of evidence is an undeniable challenge—whether the evidence involves readily quantifiable data about the spatial location of early covid cases, or qualitative observations that coronavirus research at the Wuhan Institute of Virology was conducted at relatively low laboratory bio-safety levels. And there may well be more arbitrariness in some of our weights of evidence than in others; as per the passage Bouchart highlights (Fairfield and Charman 2022, 444-45), we need to view our quantification efforts with some healthy skepticism, keeping in mind that our judgements are provisional and subject to revision. The imprecision of our weights of evidence can be partly addressed through sensitivity analysis (277, 280-82)—we specify a range of values for each piece of evidence rather than reporting only a single number. But more importantly, our estimates could serve as a starting point for structured scrutiny and debate among experts.[43]

Notwithstanding the limitation that many kinds of information do not yield objectively quantifiable probabilities, we view (approximately) objective Bayesianism as the only natural framework for knowledge accumulation, especially when it comes to learning across diverse kinds or sources of evidence. Frequentism in principle rejects the use of any data that are not generated by some stochastic process, and because probabilities cannot be assigned to theories, frequentist-based approaches are awkward at best when it comes to combining evidence or conclusions across multiple studies.[44] Fully subjective Bayesianism allows supposedly rational agents who have exactly the same information to come to different probabilistic conclusions, with no way to reconcile the discrepancy, so it is not even clear what knowledge accumulation should mean in this context. For complex and controversial cases like covid origins, the Bayesian approach offers additional benefits—it forces us to take seriously rival explanations that may run counter to what we want to be true or what we initially think is true, and it can reveal where and why reasoning about key pieces of evidence among the public, in the press, and even in peer-reviewed literature may go wrong.

Conclusion

Soifer's remarks contemplate how scholars of different persuasions will react to our work—in our experience to date, enthusiasts and skeptics have not been split along traditional quantitative vs. qualitative methodological divides. We take that as a positive sign, considering that our goal was not to write a book that everyone would agree with and readily adopt, but rather to shake up existing divisions within the discipline, rechart the methodological landscape, and challenge scholars to rethink which of their practices are justified and valuable, and which could be improved to yield more reliable and consistent inferences. We would say the more Bayes the better to that end, but to Soifer's query, we grant that readers who are reluctant to embrace the full Bayesian apparatus can still benefit from incorporating some of the lessons of Bayesian reasoning into their work. We thank the discussants again and welcome further debate moving forward.

43 Unfortunately, we have found that this particular question has become so polarized and politicized that few experts have been willing to engage in this fashion. We would also like to note here that while some political scientists have expressed trepidation about quantifying degrees of belief when working with qualitative evidence, in our view, the benefits for promoting consistency of reasoning across multiple pieces of evidence and systematically aggregating their inferential import can outweigh concerns about false precision—particularly in contexts where the evidentiary observations do not all tilt the balance in favor of the same hypothesis (Fairfield and Charman 2022, chap. 4). In these situations, drawing conclusions requires going beyond qualitative judgements about individual weights of evidence. We will have to ask, for example, whether two pieces of evidence that each moderately favor H_1 over H_2 together outweigh one piece of evidence that strongly favors H_2 over H_1. In making a judgement, we are at least implicitly moving toward quantification, and explicitly quantifying makes our decisions more transparent. If desired, one could always translate the aggregate quantified weight of evidence back into a qualitative description (e.g., weak, moderate, strong, very strong...) to avoid conveying false precision.

44 See Fairfield and Charman 2022, chap. 8.

References

Chan, Alina, and Matt Ridley. 2021. *Viral: The Search for the Origin of Covid-19*. Harper Collins.

Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing." *Political Analysis* 25 no. 3 (July): 363–80.

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

Humphreys, Macartan, and Alan M. Jacobs. 2023. *Integrated Inferences: Causal Models for Qualitative and Mixed-Method Inference*. New York: Cambridge University Press.

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

Ragin, Charles. 1997. "Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research." *Comparative Social Research* 16:27–42.

Slater, Dan. 2009. "Revolutions, Crackdowns, and Quiescence: Communal Elites and Democratic Mobilization in Southeast Asia." *American Journal of Sociology* 115(1):203–254.

Qualitative & Multi-Method Research 18(2), "Symposium: Authors meet Critics—Veil of ignorance Process Tracing."

Qualitative & Multi-Method Research 14(1)), "Symposium: Causal Mechanisms and Process Tracing."

QMMR Qualitative & Multi-Method Research

Notes from the Field

Qualitative and Multi-Method Research | Fall 2022, Vol. 20.2 | https://doi.org/10.5281/zenodo.8326524

Boundaries Unsettled: Invisible Threats and Activist Scholarship in Uruguay

Francesca Lessa

University College London

February 10, 2017, was a warm summer evening in Mendoza, Argentina. The thick blackout curtains were trying, unsuccessfully, to keep the torrid heat out of the room. In the sunset light, I glanced at my phone on the bedside cabinet. A message from my friend Silvia flashed on the screen.

Although my memories of that hot summer evening are fuzzy in places, I will never forget the content of that WhatsApp message: Silvia wanted me to know that she had heard my name on the evening news in Montevideo, Uruguay, as integrating a death list composed of 13 people, mostly authorities (including the country's attorney general and minister of defense), lawyers and human rights defenders, 10 of whom were Uruguayans and three foreigners. I knew many of them personally given the research I had been conducting on impunity for dictatorship-era crimes in Uruguay for almost ten years.

For the next few hours, I was in a shock-like state trying to make sense of what was unfolding.

Me? On a death list? In Uruguay?

I did not tell anyone about the death threats for the first 24 hours: I was unable to find the words to articulate the situation, which seemed rather surreal in those initial moments. Nothing in all the training courses I had completed as a researcher in my years at the University of Oxford—on fieldwork security, risk assessment, ethics, and vicarious trauma—could have prepared me for this.

A previously unknown group in Uruguay had disseminated the death list to the media, local authorities, and also emailed it directly some of the threatened people themselves. I had not received anything, though, aside from Silvia's message. The death threats came from the self-proclaimed "Comando General Pedro Barneix" and read as follows (IACHR 2017):

"The suicide of General Pedro Barneix will not remain unpunished… No more suicides or unjust prosecutions will be accepted. From now on, for every suicide we will kill three people selected at random from the following list."