applications. As to Boolean logit, my own experience is that its information requirements and the convoluted likelihood functions that it produces are its Achilles' heel: the former can result in inestimable models, while the latter result in frequent violations of the Wald assumptions that underpin estimated standard errors. The remedy in the first case is more data. The remedy in the second is bootstrapped standard errors, which are time-consuming but produce the correct standard error estimates.

All in all, the main critiques of these models, unsurprisingly, revolve around their position on the spectrum from high-assumption to high-information. In the case of fs/QCA, if the many assumptions fit, you can believe the results; in the case of Boolean logit, if you have enough information you can estimate the model. The two models in between, while less ambitious in terms of their ability to model causal complexity, are also less ambitious in their requirements.

Conclusion: Caveat Emptor

Taken as a whole, to reiterate, these models represent points on a spectrum, from the assumption-laden fs/QCA procedure to the data-hungry Boolean logit. The implications of data-intensivity are fairly straightforward: a data-hungry procedure runs the risk of providing null results if too few data are available to estimate all of the necessary quantities. What are the implications of incorrect assumptions?

In the case of the three statistical procedures, the main assumption has to do with the form of the interaction. A bad assumption at this stage will, in a nutshell, produce conclusions that are inaccurate to an unknowable degree—and that is every bit as bad as it sounds.

In fs/QCA, because conclusions depend on a wider range of assumptions, the cumulative implications of violating those assumptions can be even more dire. If fuzzy-set membership is estimated improperly, if the mean rather than the minimum defines joint membership in the conjunction of two sets, if the true threshold between possible and impossible cases is really $Y=X^2$, and if an independent variable's contribution to an outcome is partial (or, worse, unconditional), the results can bear shockingly little resemblance to the reality they are meant to capture.

In all cases it pays to question assumptions and to do so thoroughly. For statistical models, it is at least possible to use model fit to adjudicate among competing assumptions. We may never arrive at the One True Specification, but we can at least know which is the best given the data we have at hand. fs/QCA offers fewer assurances of this nature, but practitioners can at least get a sense of the range of possible conclusions by varying the assumptions at each step and exploring the extent to which the results are robust to those changes.

References

Berry, William D., Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54 (1): 248–266.

Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11 (3): 209–233.

Braumoeller, Bear F. and Austin Carson. 2011. "Political Irrelevance, Democracy, and the Limits of Militarized Conflict." *Journal of Conflict Resolution* 55 (2): 292–320.

Braumoeller, Bear F. and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44 (4): 844–858.

Clark, William Roberts, Michael J. Gilligan, and Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14 (3): 311–331.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30 (2): 127–145.

Hildebrand, David K., James D. Liang, and Howard Rosenthal. 1976. "Prediction Analysis in Political Research." *American Political Science Review* 70 (2): 509–535.

Hug, Simon. 2013. "Qualitative Comparative Analysis: How Inductive Use and Measurement Error Lead to Problematic Inference." *Political Analysis* 21 (2): 252–265.

Ragin, Charles C. 2013. "QCA versus Statistical Interaction." Department of Sociology, University of California, Irvine.

Seawright, Jason. 2013. "Warrantable and Unwarranted Methods: The Case of QCA." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Process Tracing with Bayes: Moving Beyond the Criteria of Necessity and Sufficiency

https://doi.org/10.5281/zenodo.894552

Andrew Bennett
Georgetown University
bennetta@georgetown.edu

Social scientists have long recognized the study of evidence from within individual cases as a fundamental tool for causal inference. This evidence helps guard against the inferential errors that can arise from making causal inferences based only on comparisons among cases. Process tracing, the systematic study of evidence from within a single case to assess alternative explanations of that case, is a key method of within-case analysis.

Yet until recently, formal articulation of the underlying logic of process tracing has been incomplete. One line of inquiry has sought to organize the traditional process tracing tests in terms of whether they provide necessary and/or sufficient grounds for inferring that a given piece of evidence confirms a particular hypothesis (Bennett 2010; Collier 2011). Thus, (1) the results of a straw in the wind test may provide suggestive, but far from definitive, support for the hypothesis; (2) the hoop test must be passed for the hypothesis to be seriously

Author's Note: This is an abridged and revised version of "Disciplining our Conjectures: Systematizing Process Tracing with Bayesian Analysis," the technical appendix to Andrew Bennett and Jeffrey Checkel, eds., *Process Tracing: From Metaphor to Analytic Tool* (Cambridge University Press, forthcoming 2014). I would like to thank Derek Beach, Jeff Checkel, David Collier, Colin Elman, Dimitri Gallow, Macartan Humphreys, Alan Jacobs, James Mahoney, Ingo Rohlfing, and David Waldner for their insightful comments on an earlier draft of this paper. Any remaining errors are my own.

entertained, and passing is therefore necessary for sustaining the hypothesis; (3) the smoking gun test affirms the hypothesis and passing is therefore sufficient for sustaining the hypothesis, although it does not exclude other explanations, and (4) a doubly decisive test affirms the hypothesis and excludes other explanations. This framework draws informally on Bayesian logic, but strictly speaking, Bayesianism requires that we never be one hundred percent convinced of the truth or falsity of any explanation, so the terms "necessary" and "sufficient" are too categorical.

Bayesianism is not the only way to understand process tracing. James Mahoney (2012) has demonstrated that set theory can be used to arrive at many of the same insights regarding process tracing, provided that no evidence is considered fully necessary or sufficient to judge explanations as either certain or impossible. Similarly, David Waldner (forthcoming) has argued that directed acyclic graphs are a useful way to think about process tracing.[1] Process tracing has been most fully explicated in terms of Bayesianism, however, and the following discussion continues this approach.[2] It concludes that using Bayesian logic more fully, systematically, and transparently can improve the quality and replicability of process tracing and strengthen causal inferences, including those based on qualitative and quantitative cross-case comparisons.[3]

Fundamentals of Bayesian Analysis

Process tracing analyzes within-case evidence to develop or test explanations of individual cases. Doctors diagnosing patients, detectives investigating crimes, and social scientists developing and testing both general theories and historical explanations of particular cases are all interested in how we should update theories and explanations in the light of evidence from individual cases. One of the most powerful ways of thinking about this challenge is the logic first systematized by Thomas Bayes in the mid-1700s. Bayes focused on the question of how we should update our confidence in an explanation given new, relevant evidence. This updated confidence in the likely truth of a theory is referred to as the posterior, or the likelihood of a theory conditional on the evidence.

In Bayes's approach, we need three key pieces of information, in addition to the evidence itself, to calculate this posterior likelihood. First, we need to start with a "prior" likelihood, which expresses our initial confidence that a theory is true even before looking at the new evidence. For example, let us assume we have an explanation of a case that we think is 40 percent likely to be true, and for simplicity let us further assume that it is mutually exclusive with the alternative explanations—that is, only one could be true—so the likelihood it is false is one minus 40 percent, or 60 percent.[4]

Second, we need information on the likelihood that, if a theory is true in a given case, we will find a particular kind of evidence for that case. This is referred to as the evidence conditional on the theory. We can view the theory as an attempt to capture the underlying "data generating process," and hence as a useful way to understand the claim that the evidence is conditional on the theory. Let us assign this a likelihood of 20 percent to illustrate a "smoking gun" test. This is a test in which confirmatory evidence, if found, strongly increases our confidence in the explanation, but the failure to find that evidence does not strongly undermine our confidence in the explanation.

Third, we need to know the likelihood that we would find the same evidence even if the explanation of interest is false—i.e., a false positive. In our example, to complete the logic of a smoking gun test, let us assign this a probability of 5 percent.[5]

Smoking Gun Test

Analysis of the three estimated probabilities necessary for Bayesian updating of our explanation can be illustrated with the smoking gun test. Using P for the explanation, *pr*(P) for the prior probability that P is true, and k for the evidence, we have:

[1] It is not yet clear whether there are methodologically consequential differences among Bayesianism, set theory, flow graphs, and directed acyclic graphs with regard to process tracing. There are many ways in which these three logics are compatible and translatable; on this point, see Zdislaw Pawlak, "Bayes' Theorem—the Rough Set Perspective," at http://bcpw.bg.pw.edu.pl/Content/1935/btrsp_or.pdf, accessed May 1, 2014, and Abell (2009: 45–58).

[2] Bennett (2008); Abell (2009); Beach and Pedersen (2013a, 2013b); Collier (2011); Humphreys and Jacobs (2013); Mahoney (2012); and Rohlfing (2012, 2013a, 2013b).

[3] This point has often been made with regard to combining statistical analysis and within-case analysis. However, Bayesian analysis can also strengthen qualitative methods of cross-case comparisons, whether typological theory (George and Bennett, 2005) or Qualitative Comparative Analysis (QCA) (Ragin 2008). These qualitative methods are greatly strengthened by combining them with process tracing. On the latter point, see Schneider and Rohlfing (2013).

[4] One complication is that theories or explanations may not be mutually exclusive, but rather complementary. If I sneeze, for example, it may be due to allergies, to having a cold, to sudden exposure to bright lights, or to a combination of any two or all three factors; thus, showing that there was exposure to bright light does not necessarily raise or lower the likelihood that having a cold or allergies contributed to my sneezing. The present discussion, like many pedagogical presentations of Bayesianism, simplifies this point by considering only whether one explanation is true or false, and assuming other theories are mutually exclusive, so the likelihood that the explanation is false is one minus the likelihood that it is true (see also Rohlfing 2012: chap. 8). In social science research, researchers often face the more complex question of hypotheses that, overall, are partly complementary and partly competing; or, alternatively, competing in the context of some cases and complementary in others (on this challenge see Rohlfing 2013a).

[5] Ideally estimates of priors and of the likelihood of finding evidence depending on whether a theory is true or false would be based on studies of many prior cases or well-validated theories or experiments. This is true in the medical research examples common in textbook discussions of Bayesianism. Unfortunately, in the social sciences we often lack such data and must begin with more subjective guesses on these probabilities. The reliance on subjective expectations of probabilities, and differences in individuals' estimates of these probabilities, is an important challenge for Bayesianism, although strongly probative evidence can lead to convergence between observers who start with greatly different assumptions on their priors.

Smoking Gun Test

Prior likelihood P is true, or $pr(P) = .40$
Likelihood of smoking gun evidence k, if P is true = .20
Likelihood of smoking gun evidence k, if P is false = .05

We can now address the following question: if the evidence supporting the explanation is found, what is the updated likelihood that the explanation is true?

In a common form of Bayes' Theorem, the updated likelihood that a proposition P is true in light of evidence k, or $Pr(P|k)$, is as follows:

$$Pr(P|k) = \frac{pr(P)pr(k|P)}{pr(P)pr(k|P) + pr(\sim P)pr(k|\sim P)} \quad (1)$$

Notation:
$Pr(P|k)$ is the posterior or updated likelihood of P given (i.e., conditional on) evidence k
$pr(P)$ is the prior likelihood that proposition P is true
$pr(k|P)$ is the likelihood of evidence k if P is true (or conditional on P)
$pr(\sim P)$ is the prior likelihood that proposition P is false
$pr(k|\sim P)$ is the likelihood of evidence k if proposition P is false (or conditional on ~P)

If we put our illustrative numbers into equation (1), the updated likelihood of the explanation being true is .73:

Likelihood the explanation is True for a Passed Smoking Gun Test

$$\frac{(.4)(.2)}{(.4)(.2) + (.6)(.05)} = \frac{.08}{.08 + .03} = \frac{.08}{.11} = .73 \quad (2)$$

We can use Bayes' theorem to calculate the posterior likelihood of a failed smoking gun test to be .36. Hence, as the name of the test implies, passing the test raises the theory's likelihood far more (from .4 to .73) than failing it would lower this likelihood (from .4 to .36). This illustrates a key feature of Bayesianism. The extent of updating when a test result is positive is driven by the prior likelihood of the theory and the likelihood ratio, which is the ratio of true positives to false positives (Rohlfing 2013b).[6] Here, the likelihood ratio for positive evidence on the smoking gun test is:

$$\frac{\text{Likelihood of true positive}}{\text{Likelihood of false positive}} = \frac{.2}{.05} = 4 \quad (3)$$

The higher the likelihood ratio (above a minimum value of 1) the more powerful or discriminating the evidence: finding positive evidence when the likelihood ratio was 4, as in the smoking gun test example, greatly increases the likelihood that the proposition is true.[7] When the likelihood ratio is equal to one, evidence has no discriminatory power: the posterior is the same as the prior.

Straw in the Wind, Hoop and Doubly Decisive Tests

The other three tests also exhibit continuous gradations in their strength. Hoop tests are the converse of smoking gun tests. In a hoop test, the absence of confirming evidence strongly undermines an explanation, but the presence of such evidence does not strongly increase the likelihood that the explanation is true. A straw in the wind test provides only weak evidence for or against an explanation. Finally, a doubly decisive test strongly increases the likelihood of an explanation that passes, and strongly undermines that of an explanation that fails.

Macartan Humphreys and Alan Jacobs (2013) have devised an excellent diagrammatic representation of how the likelihood ratio establishes the strength of these evidentiary tests. Figure 1 (adapted from Humphreys and Jacobs 2013:17), shows how these tests relate to the two measures that comprise the likelihood ratio: the likelihood of observing evidence k when a proposition P is true (labeled q_1 on the y-axis of the figure) and the likelihood of observing evidence k even when the proposition P is false (labeled q_0 on the x-axis of the figure).

The figure brings into sharp focus the mirror-image relations among tests depending on whether evidence k is present or absent. A test that provides smoking gun evidence for P when k is present constitutes hoop test evidence for ~P when k is absent, and vice-versa. Similarly, a hoop test for P is a smoking gun test for ~P. This is because P and ~P are inversely proportional—their probabilities add to one.

Humphreys and Jacobs also introduce a set of figures that further illustrate the properties of different evidentiary tests, again reproduced here as Figures 2 to 5.[8] These figures show how different prior probabilities map onto posterior probabilities for the illustrative likelihood ratio used in each graph. Examples are shown for likelihood ratios representing hoop, smoking gun, doubly decisive, and straw in the wind tests. Because q_0 and q_1 can vary continuously between zero and one, in addition to the examples in Figures 2 to 5, one could draw any number of curves for tests of different discriminatory power within each family of tests.[9]

These graphs nicely illustrate the point that the extent to which we should update our prior depends on the values of both the prior and the likelihood ratio. As Humphreys and Jacobs point out, we will not lose as much confidence in a hypothesis that has achieved a high prior through repeated earlier testing, even in the face of a failed hoop test. In Figure

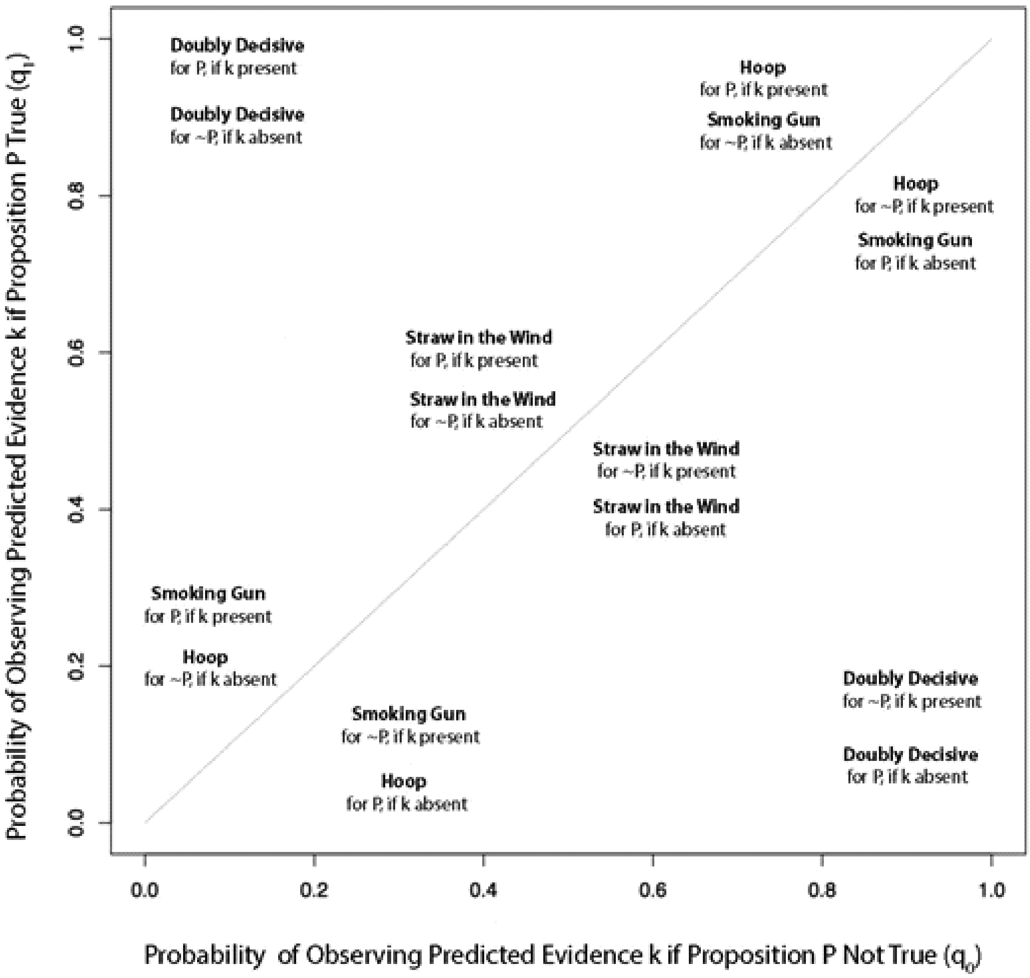
[6] For arguments that the likelihood ratio, or more specifically the log of the likelihood ratio, is the best measure of the evidential or confirmatory support of evidence, see Fitelson (2001) and Eels and Fitelson (2002).

[7] There is also a likelihood ratio with regard to a negative finding. This is the ratio of the likelihood of a false negative divided by that of a true negative. This ratio ranges from zero to one, and the closer it is to zero, the more powerful a negative finding is in undermining the likelihood that an explanation is true. The likelihood ratio for a positive finding is designated as LR+, while that for a negative finding is designated LR-. For present purposes, I use the term "likelihood ratio" to refer to LR+.

[8] Humphreys and Jacobs (2013: 19); see also Rohlfing (2013b: 20–29).

[9] On this point, see Mahoney (2012) and Rohlfing (2013a).

Figure 1: Mapping Process Tracing Tests Based on the Likelihood Ratio



Note: Figure adapted from Humphreys and Jacobs (2013: 17), with permission of the authors. P = Proposition being tested. k = Evidence evaluated to carry out test. q_0 and q_1 = Probability of finding evidence k, according to falsity or truth of proposition P

3 the vertical distance from the 45-degree diagonal to the curved line for the failed hoop test, which shows how much lower the posterior is than the prior, is less when the prior is close to 1.0 than when it is when the prior is between 0.4 and 0.8.

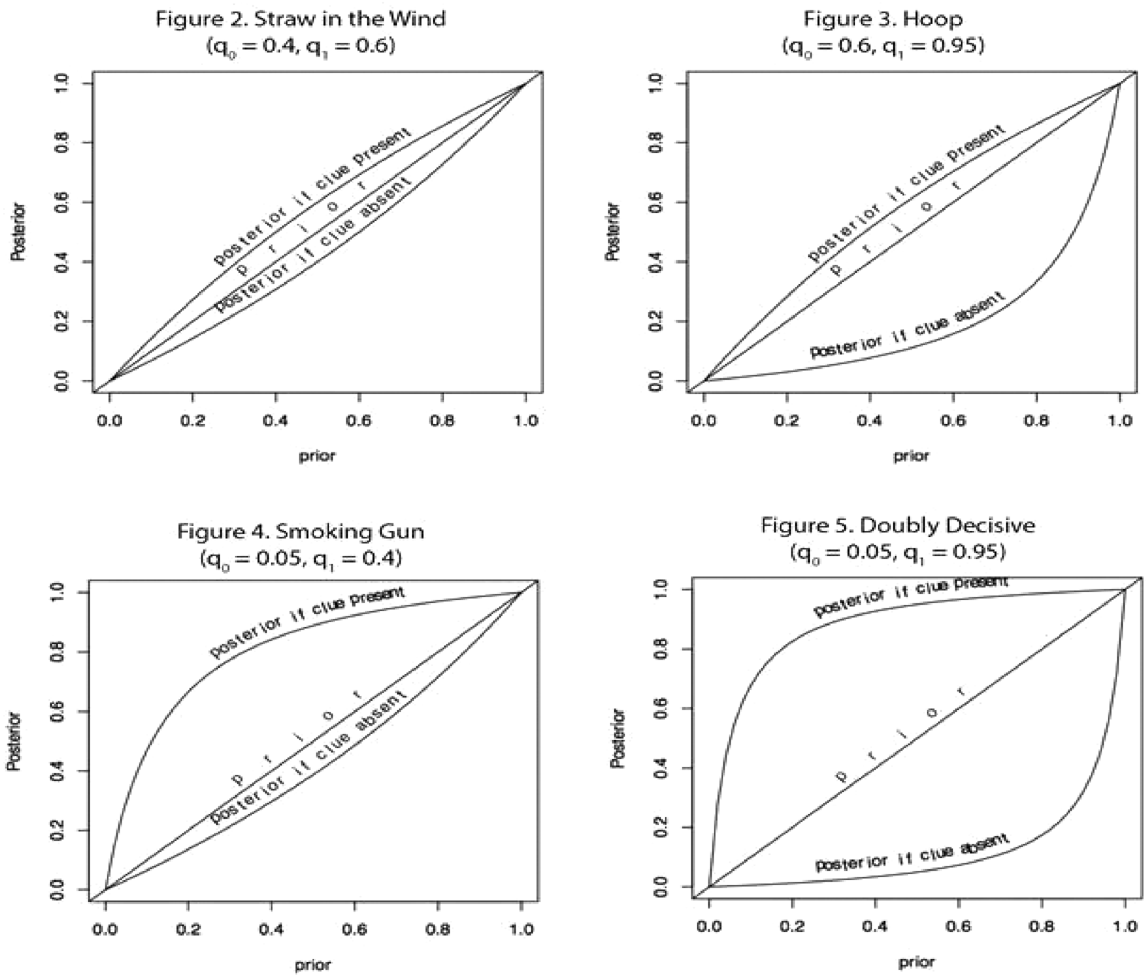
The mathematical relationships among q_0, q_1, the prior, and the posterior allow us to test for consistency. Given any three of these likelihoods, we can determine what the value of the fourth must be if our thinking is consistent. Alternatively, given any two, we can determine what the ratio of the other two should be. For example, given an individual's prior and posterior, we can determine what their likelihood ratio should have been for the evidence they examined. In cases where a scholar has a prior of 40 percent and a posterior of 95 percent, we know that their likelihood ratio for the evidence they examined should have been just over 28. That is, they should have been 28 times as likely to expect the evidence if the theory was true than if it was false, which is an extremely high ratio. If the scholar did not think this likelihood ratio was justified, they might have to lower their estimate of the posterior.

The Implications of Bayesianism for Process Tracing

I explore elsewhere how Bayesian logic reveals a number of implications for process tracing (Bennett 2008; Bennett forthcoming 2014); here, I focus on five.

First, the explication of Baysianism above improves upon my earlier writings on the subject. Earlier, I infelicitously suggested that it was *necessary* for an explanation to pass a hoop test in order to remain viable, whereas passing a smoking gun test was *sufficient* to confirm an explanation (Bennett 2010). This language is misleading, in that Bayesianism reminds us that we can never be 100 percent confident that an explanation is true, or that it is false. There are several reasons for this. First, there may be alternative explanations that fit the evidence better. Second, there is always some evidence that is inaccessible. Third, there may be errors in the measurement of the evidence. More generally, we cannot tell for certain if a theory's failure in an evidentiary test undermines the theory or if it undermines auxiliary hypotheses, explicit or implicit in the theory, about the observation and measurement of evidence.

Figures 2 to 5: Illustrative Examples of the Four Tests
(Adapted from Humphreys and Jacobs, 2013: 19.)



Thus, although some pieces of evidence may be highly probative, we cannot infer with certainty that a theory or explanation is true based on the evidence.

Second, counter-intuitively, evidence consistent with a theory can actually lower its posterior because this same evidence is even more consistent with an alternative theory. Conversely, evidence that does not fit a theory can actually raise its posterior by more severely undermining an alternative explanation. These outcomes happen when the likelihood ratio is less than one.[10] Figures 2 to 5 all have likelihood ratios where q_1 is greater than q_0; that is, they are all drawn from above the 45-degree diagonal in Figure 1. When q_0 is greater than q_1, the likelihood ratio is less than one (as in the area below the 45-degree diagonal of Figure 1), and evidence consistent with P actually reduces the likelihood that P is true.[11]

Third, Bayesianism provides a logical rationale for the methodological prescription that independence and diversity of evidence is important in process tracing. Desirable evidentiary tests are those that are independent of one another, and diverse—i.e. they bear on different alternative hypotheses. Regarding independence, if one piece of evidence is wholly determined by another, it has zero additional power to update prior probabilities. As for diversity of evidence, as we accumulate more and more pieces of evidence that bear on only one alternative explanation, each new piece has less power to update further our confidence in that explanation. This is true, even if the evidentiary tests are independent, because we have already incorporated the information of the earlier, similar evidence.

Fourth, multiple weak tests, if independent from one another, can sometimes cumulate to strongly update priors. Straw in the wind tests, and weak smoking gun and hoop tests, are the kinds of tests that might be called "circumstantial evi-

[10] See also Rohlfing (2013b: 5, 19, 20).

[11] In medical tests, the positive likelihood ratio as discussed in footnote 7 above, LR+, is simply *defined* as the test result that makes it more likely a patient has a particular disease. If a doctor thought a certain test result was likely to be associated with the disease, but found the opposite to be true, she or he would simply flip the interpretation of what reading on the test constituted a "positive" outcome. With the testing of posited social mechanisms, however, social scientists do not necessarily flip the interpretation of what it means to find that the hypothesized evidence of the mechanism was observed.

dence" in a court case. If most of these kinds of tests point in the same direction, this provides strong evidence for the explanation in question. This is analogous to the high likelihood that a coin is biased toward heads if it comes up heads significantly more than 50 percent of the time in a large number of fair coin tosses.

The final implication points to a crucial choice: whether to "fill in the numbers" by explicitly assigning priors and likelihood ratios and using Bayesian mathematics, at least for the few pieces of evidence that a researcher considers the most probative, in an effort to make process tracing more rigorous and transparent. Earlier discussions treated Bayesianism as a useful metaphor for process tracing (McKeown 1999) or a way of clarifying its logic (Bennett 2008), without arguing that Bayesian mathematics should be used explicitly in process tracing. Other researchers also argue that more explicit use of Bayesian mathematics in process tracing is impractical and would convey a false sense of precision (Beach and Pedersen 2013a).[12] More recently, however, a number of scholars (Abell 2009; Humphreys and Jacobs 2013; Rohlfing 2013b) have suggested that researchers should in fact implement Bayesianism more concretely, explicitly identifying their priors and likelihood ratios and using Bayes' theorem to determine posterior probabilities.

A powerful argument for actually filling in the numbers in process tracing is that it asks researchers to make specific and transparent the assumptions that they must in any case make implicitly if process tracing is to have probative value. The process of clearly identifying the likelihood of finding a certain kind of evidence, not only conditional on the truth of a theory but also conditional on the falsity of the theory, can push researchers to clarify their own thinking. It also makes this thinking more transparent to other scholars, eliminating the considerable ambiguity in many verbal formulations used to convey the likelihoods of explanations and evidence.

We have good examples of process tracing in which scholars have been exceptionally careful and explicit in the evidence they used and the type of tests (e.g. hoop tests, smoking gun tests) they applied in making inferences (Fairfield 2013). So far, however, we have no full-fledged examples where scholars have done process tracing with explicit priors and numerical Bayesian updating; this remains an area where the advice of at least some methodologists diverges from the practices of working researchers.[13] Whether one ultimately prefers to use Bayesian logic implicitly or explicitly, understanding this logic unquestionably helps clarify the logic of process tracing.

References

Abell, Peter. 2009. "A Case for Cases: Comparative Narratives in Sociological Research." *Sociological Methods and Research* 38 (1): 38–70.

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

Beach, Derek and Rasmus Brun Pedersen. 2013b. "Turning Observations into Evidence: Using Bayesian Logic to Evaluate What Inferences are Possible Without Evidence." Presented at the Annual Meeting of the American Political Science Association, Chicago.

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

Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed., eds. Henry E. Brady and David Collier. Lanham, MD: Rowman & Littlefield, 207–219.

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

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

Eels, Ellery, and Branden Fitelson. 2002. "Symmetries and Asymmetries in Evidential Support." *Philosophical Studies* 107 (2): 129–142.

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

Fitelson, Branden. 2001. "A Bayesian Account of Independent Evidence with Applications." *Philosophy of Science* 68 (3): S123–S140.

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

Humphreys, Macartan, and Alan Jacobs. 2013. "Mixing Methods: A Bayesian Unification of Qualitative and Quantitative Approaches." Presented at the Annual Meeting of the American Political Science Association, Chicago.

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

McKeown, Timothy. 1999. "Case Studies and the Statistical World View." *International Organization* 53 (1): 161–190.

Pawlak, Zdislaw. "Bayes' Theorem—the Rough Set Perspective." http://bcpw.bg.pw.edu.pl/Content/1935/btrsp_or.pdf.

Ragin, Charles. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. New York: Palgrave Macmillan.

Rohlfing, Ingo. 2013a. "Comparative Hypothesis Testing via Process Tracing." *Sociological Methods and Research*. Advance online publication. doi: 10.1177/0049124113503142

Rohlfing, Ingo. 2013b. "Bayesian Causal Inference in Process Tracing: The Importance of Being Probably Wrong." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Schneider, Carsten and Ingo Rohlfing. 2013. "Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research." *Sociological Methods and Research* 42 (4): 559–597.

Waldner, David. Forthcoming 2014. "What Makes Process Tracing Good? Causal Mechanisms, Causal Inference, and the Complete ness Standard in Comparative Politics." In *Process Tracing*, eds. Andrew Bennett and Jeffrey Checkel. (Cambridge: Cambridge University Press).

[12] But see Beach and Pedersen (2013b) which urges more explicit and transparent use of Bayesian logic, if not specific use of mathematical probability estimates.

[13] Abell (2009: 59–61) provides a brief illustrative example of explicit Bayesian updating in process tracing. In this example, he uses a panel of trained researchers, rather than an individual researcher, to estimate likelihood ratios based on shared evidence from the case.