Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44, no. 4 (October): 823–30.

Copestake, James, Marlies Morsink, and Fiona Remnant, eds. 2019. *Attributing Development impact: The Qualitative Impact Protocol Case Book*. Rugby: Practical Action Publishing. https://practicalactionpublishing.com/book/105/attributing-development-impact.

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): 154-167.

Goertz, Gary, and Stephan Haggard. 2021. "Generalization, Case Studies, and Within-Case Causal Inference: Large-N Qualitative Analysis (LNQA)." In H. Kincaid and J. Van Bouwel (eds.) *The Oxford handbook on the philosophy of political science*. Oxford: Oxford University Press.

Levitsky, Steven, and Lucan Way. 2010. *Competitive Authoritarianism: Hybrid Regimes After the Cold War*. Cambridge: Cambridge University Press.

Mahoney, James. 2007. "The Elaboration Model and Necessary Causes." In *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, edited by Gary Goertz and Jack Levy, 281-306. London: Routledge.

___. Forthcoming. *The Logic of Social Science*. Princeton: Princeton University Press.

Seawright, Jason, and John Gerring, 2008. "Case selection techniques in case study research: a menu of qualitative and quantitative options." *Political Research Quarterly*, no. 61: 294-308.

Waldner, David. 2015. "Process Tracing and Qualitative Causal Inference." *Security Studies* 24, no. 2 (June): 239–50.

VPT: A Bad Idea Whose Time Has Come?

Derek Beach
Aarhus University

This has been a difficult review to write. On the one hand, it is important to be constructive and open to new approaches for how to do good process tracing (PT). On the other hand, the proposed "Veil of ignorance Process Tracing" (VPT) attempts to solve one vaguely defined problem through a confused mismatch of ideas that result in a set of much bigger problems. Building on my experiences both as a practitioner of multiple PT case studies (Beach, Schäfer, and Smeets 2019; Beach and Smeets 2020), and a methodologist who has published whole chapters detailing how to guard against confirmation bias (see e.g., Beach and Pedersen 2019), my diagnosis of VPT is that there are a series of highly problematic (and puzzling) theoretical, methodological, and practical issues that make for a recipe producing poorly-evidenced empirical generalizations as to why an event took place that will not result in a real contribution to our broader understanding of how causal processes work in real-world cases.

The authors postulate—without evidence—that confirmation bias is an endemic problem in process tracing, insinuating that much existing PT consists of "just-so stories" where researchers search for evidence that confirms their pet theory. In reality, contrary to what these authors write, confirmation bias in non-experimental science is not avoided by "blinding," but instead by being one's own strongest critic when evaluating empirical evidence, achieved in PT through thorough and transparent assessment of the uniqueness of found evidence in relation to alternative explanations (see e.g. Fairfield and Charman 2017; Beach and Pedersen 2019). Further, good peer review should detect this bias—but is not mentioned by the authors.

Before I detail these problems, a background note is important to understand the context within which VPT was developed. The method comes from the field of policy evaluation, within which there is a widespread belief amongst *funders* that the only scientific method that avoids confirmation bias in evaluations is to use counter-

factual-based assessments in the form of experimental methods (termed randomized controlled trials (RCTs) in evaluation).[1] The VPT method reads as if it were originally framed to allay funder concerns about evaluation case studies by including a number of buzzwords from experimentalist research. But this begs the question: Why should political science or international relations adopt this unhappy marriage between a naïve, empiricist version of grounded theory with ideas about blinding research processes that are appropriate when using RCTs?

Confused Assertions About Types of Causal Claims Being Made

VPT asserts itself as a method for making causal attributions, but, in my opinion, it is all over the place regarding the *nature* of the causal claims being assessed. The proposal uses the terminology of causal effects throughout but also proposes eight (!) different types of causal relationships that can be explored using VPT. These include [positively/negatively/not related] as a [necessary/sufficient] cause [mediated/interacted with Z] that had a [weak/strong] impact on Y, potentially with [positive/negative] feedback. At the same time, the proposal insists that: *"The veiling methodology helps identify elements of a causal mechanism"* (this issue, italics in original).[2] As its name suggests, PT is all about tracing causal processes (aka mechanisms) that link a cause (or set of causes) and an outcome together (e.g., Bennett and Checkel 2014; Schmitt and Beach 2015; Beach and Pedersen 2019). This also implies that PT—as a within-case method—cannot assess net causal effects because of the problem of masking.[3]

At no point in the article is the *causal* nature of mechanistic explanations developed (see below), which together with the eight different types of causal claims that the VoiPT supposedly can evidence, results in complete confusion about the ontological properties of "causal claim Lego blocks."[4] The authors even readily admit this on the first page, writing: "Our approach starts with the proposition that diverse forms of qualitative evidence… all contain a *variety* of causal claims (this issue, italics added)." As a result, the VPT proposal forges ahead by asserting that "causal claim Lego blocks" are just lying in wait in the empirical record, like causal gold nuggets that the research assistant (RA) can snatch up.

Process Tracing Without the Mechanisms

If we put aside the problems created by the numerous types of causal claims involved, the VPT proposes that the found causal gold nuggets are compared by the principal researcher (PR) with her "favored" causal mechanism in the "integration" stage of VPT. However, what this involves is unclear because the VPT appears to understand mechanisms as a more descriptive series of events. The article uses the example of Levy's work on the outbreak of continental war (WWI). When we look at the series of events put forward by Levy in the figure, however, it quickly becomes obvious that causation is being *assumed* through temporal sequencing of events, but there is no causal theorization of the linkages between them. For example, there is no explanation for why the combination of "Russian beliefs" and "Austria loses some legitimacy" produces the next step, "Russian mobilization." What is the causal linkage between a static factor (Russian beliefs) and an unexplained event (Austria losing some legitimacy)? Therefore, the explanation produced is a series of factors/events, not a mechanistic causal explanation.

But how then can *causal* gold nuggets supplied by the RA aimed at explaining *why* the outcome occurred in the case be compared with a more *descriptive* series of events (i.e. *what* happened) that the PR has developed beforehand? And even if we ignore this mismatch, the set of causal gold nuggets delivered by the RA would only tell the PR that a majority of stakeholders believed that X was causally related in one or more of the eight types of causal types (see above). How a plausible causal mechanism could be reconstructed on the basis of such a disparate set of causal gold nuggets about why an outcome or event occurred is difficult to see.

Lack of Alignment Between Type of Causal Claim and Methods to Make Inferences

Moving to the issue of inference, as we know from Hall's now famous principle of methodological alignment (2003), our methodology has to be appropriate

---

1 Interestingly, while many funders hold these beliefs, many methodologists in the field assert that other methods (e.g., contribution analysis) are just as robust but importantly that they also answer different questions with different types of evidence (e.g. Scriven 2008; Mayne 2012; Cartwright 2012; Schmitt and Beach 2015).

2 One of the eight types of "causal claims" that is not even causal in nature. "Certainty about the causal claim" is an epistemological question related to the strength of the evidence backing a causal claim.

3 It is now well established in the philosophy of science that tracing a causal mechanism in a case does *not* enable assessment of net causal effects, only that there is a mechanism linking them. See Clarke et al. 2014.

4 The VPT proposal also does not get Lego terminology correct. Lego has *bricks*, not blocks. See www.lego.com.

for evidencing the *type* of causal claim we are making. Alignment means that evidencing a claim about *necessity* requires a counterfactual comparison (either using actual or hypothetical cases), whereas evidencing a claim about the *strength* of a causal relationship requires quantitative cross-case data about variation in X, Y, and confounders, ideally interval-scale data that can be assessed using multiple regression or other appropriate statistical tools to figure out the strength of the net causal effect. As regards a claim about necessity, is the VPT suggesting that when the RA asks a stakeholder about why event Y occurred, and they respond by saying that they think event X was necessary for Y to occur, that this is a valid causal inference? Can we assume that the stakeholder engaged in a proper counterfactual, cross-case assessment of the claim (ideally from a natural experiment)? Or are we just accepting the causal attribution made by the stakeholder as such, in effect accepting *hearsay* as evidence confirming the claim about necessity?

Induction and the Myth of Data Speaking for Itself

Another fundamental issue relates to the issue of inferences and moving from empirics back to theory. The VPT proposal builds on a naïve empiricist reading of grounded theory, where theories are built *inductively* from observable data. In effect, the VPT assumes that causal gold nuggets are just lying in wait in an uncontaminated state, where it is the ignorance of theory and empirics that enables the RA to extract these "causal claims from the data" without being tainted by prior theoretical bias. But despite developing elaborate coding procedures, the naïve empiricist reading of grounded theory itself never lived up to its promise, because data does not speak for itself. As is well established in the philosophy of science, observation of the world is *per definition* theoretically informed (Tavory and Timmermans 2014, 13-17; Jackson 2016). Causal claims are what give empirical data explanatory meaning, with scientific research as a sustained dialogue between theory and empirics. For example, when an analyst finds a puzzling rhetorical pattern in a speech by a politician, existing theories coupled with detailed empirical knowledge about what is typically in speeches are the tools that enable the analyst to figure out what the pattern means and thereby also what theoretical proposition it could be evidence of.

In my substantive research I have spent many hours reading and analyzing archival documents, during which I have *never* seen a causal gold nugget just lying around in a policy brief, intelligence assessment, or draft proposal.

Instead, working with archival information is a continuous back-and-forth between focused searches based on theoretical hunches, finding new puzzling things in documents, and then returning to theory (and case knowledge) to make sense of them, and back again. The idea that a novice RA could extract causal gold nuggets from the empirical record without any bias is therefore difficult to see.

Inherent Bias in Stakeholder Attributions of Causality

VPT suggests that causal attributions by stakeholders are *less* prone to bias. VPT proposes to use statements by stakeholders about why something happened, drawn from interviews or archival documents they have produced, supplemented with (ideally randomly selected!) attributions found in the secondary literature. The authors write that as the "source agent does not know the theory under investigation, it is hard for her to have any significant bias" (this issue).

But there is a very simple reason we use the term stakeholder—the person has a STAKE in the issue being discussed. These statements can act as evidence of what a given stakeholder wants to be heard saying for the sake of posterity. And even if they could be trusted, there can be a myriad of reasons detailed by cognitive psychologists for why their perceptions might not match the perceptions of an outside observer. This is why historians typically do *not* use stakeholder interviews, or trust stakeholder confessionals in archival documents. Instead, they want to gain access to archival documents that can enable them to reconstruct a historical process. Political scientists often do not have this luxury, and therefore have to subject these types of confessionals with robust source criticism—which raises the next point.

Inexperienced RAs Cannot Engage in Source Criticism

The suggestion of using *inexperienced* RAs to collect and code empirical material is problematic. The authors write that the RA "need not even be familiar with the empirical field in question except to the extent absolutely necessary" (this issue). However, this inexperienced RA would lack the case-related knowledge to be able to do a good interview (e.g., follow-up questions), and more critically, would not be able to engage in proper source criticism of interviews, archival documents, or secondary sources—all of which would contain implicit or explicit bias that requires significant theoretical and empirical knowledge to evaluate (see Beach and Pedersen 2019, 195-222; Møller and Skaaning 2018). Evaluating what

statements of stakeholders mean in a particular case, and whether we can trust them, requires significant case-specific knowledge and experience. Often evidence of parts of causal mechanisms are not simple "confessions," but instead involve analyzing patterns of speech, sequencing of events, or even just the fact that two social actors held a meeting. All of these empirical observations would have to be evaluated in relation to a theory *before* they could be *evidence of* anything. Good theoretical explanations cannot just be extracted by a novice from the data! If this were the case, the most compelling research would be found in the first essays done by political science students in our bachelor programs.

Therefore, the blinding of the VPT creates the very real risk that the RA regurgitates the causal gold nuggets found in statements from a majority of interviewees, archival documents, or secondary sources in the report to the PR. The PR expert is several steps removed from the nitty-gritty of the sources, preventing them from engaging in real source criticism when evaluating the evidential backing for causal gold nuggets. The result is the de facto replacement of a good, experienced researcher with a novice.

And Would I Really Send an Ignorant RA to do Interviews?

On a final practical note, deliberately sending an inexperienced RA to collect causal gold nuggets from interviews with high-level stakeholders would run into practical difficulties. In my own substantive research on high-level crisis governance in the EU, would I really send an RA who knows nothing about how Brussels works or the dossiers I am studying to conduct an interview on my behalf? One of the first things that happens in elite interviewing is that high-level civil servants test you to see whether you have done your homework (e.g., having read all of the proposals and public accounts of the negotiations). If they find you wanting, the interview will very quickly end because you are wasting their time. On the upside, once you have passed this "test," they can be very useful informants, often offering further non-public documents and drafts, along with helping you understand the signals embedded in the often-arcane language they use in legislative proposals and communications. All of this empirical material that I can use to understand how things work in Brussels would be completely out of reach for an inexperienced RA.

Conclusion

This review might seem excessively harsh. It was not intended to be, but the VPT proposal has a fractal character in which the more you dig into it, the more confused and problematic the suggestions become. The article tries to validate the method by claiming that "the idea has been field tested" in many evaluations (this issue). However, just because something has been done does not make it right, and it cannot mask the fundamental flaws in the theoretical and empirical logic inherent in the naïve empiricism of VPT, coupled with the problematic idea of blinding.

Fortunately, the state-of-the-art regarding PT methods is sufficiently developed, such that there exist robust methodological alternatives that are being widely used by researchers to produce important, evidence-based findings about causal mechanisms at play in the real world (see Beach 2020 for a review of recent uses and methodological developments). VPT is a bad idea whose time does not need to come.

References:

Beach, Derek. 2020. "Process Tracing Methods." In *Oxford Bibliographies Online—International Relations*. DOI: 10.1093/OBO/9780199743292-0227.

Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process-Tracing Methods: Foundations and Guidelines.* 2nd Edition. Ann Arbor: University of Michigan Press.

Beach, Derek, and Sandrino Smeets. 2020. "The Unseen Hands: Collaborative Instrumental Leadership in the British Renegotiation Case." *European Journal of Political Research* 59, no 2 (May): 444-464.

Beach, Derek, David Schäfer, and Sandrino Smeets. 2019. "The Past in the Present—The Role of Analogical Reasoning in Epistemic Learning About How to Tackle Complex Policy Problems." *Policy Studies Journa.* doi.org/10.1111/psj.12372.

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

Cartwright, Nancy. 2012. "Presidential Address: Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79, no 5 (December): 973-989.

Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing. Guidelines, Opportunities, and

Caveats." *Political Analysis* 25, no 3 (July): 363-380.

Hall, Peter A. 2003. "Aligning Ontology and Methodology in Comparative Politics." In *Comparative Historical Analysis in the Social Sciences*, edited by James Mahoney and Dietrich Rueschemeyer, 373–404. New York: Cambridge University Press.

Jackson, Patrick T. 2016. *The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics*. 2nd ed. London: Routledge.

Mayne John. 2012. "Contribution Analysis: Coming of Age?" *Evaluation* 18, no. 3 (July): 270-280.

Møller, Jørgen, and Svend-Erik Skaaning. 2018. "The Ulysses Principle: A Criterial Framework for Reducing Bias When Enlisting the Work of Historians." *Sociological Methods & Research* 1-32 doi.org/10.1177/0049124118769107.

Schmitt, Johannes, and Derek Beach. 2015. "The Contribution of Process-Tracing to Evaluations of Budget Support Programs." *Evaluation* 21, no. 4 (October): 429-447.

Scriven, Michael. 2008. "A Summative Evaluation of RCT Methodology & An Alternative Approach to Causal Research." *Journal of MultiDisciplinary Evaluation* 5, no. 9 (March): 11–24.

Tavory, Iddo, and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.

A Bayesian Perspective on Theory-Blind Data Collection

Tasha Fairfield
London School of Economics

Copestake, Goertz, and Haggard's (CGH) "Veil of ignorance Process Tracing" (VPT)—which in essence entails placing a firewall between data collection and hypothesis testing[1]—is an interesting addition to a growing list of proposals made in recent years that aim to address potential sources of bias in qualitative social science. Many of these proposals (e.g., pre-registration, time-logging whether evidence was discovered before or after a hypothesis was devised) import prescriptions from large-N, frequentist, statistical research that, from a Bayesian perspective, are not applicable to qualitative research. Bayesian reasoning provides its own safeguards against the problems of confirmation bias and ad hoc hypothesizing, without imposing procedural constraints that would interfere with the inherently iterative, dynamic, and interactive nature of case-study research—where we go back and forth between hypothesizing, data collection, and analysis (Fairfield and Charman 2019).

My comments begin by outlining the costs (which seem significant) and gains (limited, in my analysis) of firewalled data collection in qualitative research. I then discuss what I interpret as a fundamental shortcoming with the authors' approach that seems to undermine its core aim of separating data collection from hypothesis testing—namely, conflating evidence, evidentiary sources, and causal claims. Finally, I briefly outline my preferred approach for managing the problems of confirmation bias, ad hoc hypothesizing, and cherry-picking.

Scrutinizing the Costs and Benefits of Firewalled Data Collection

As with suggestions for pre-registration or time-logging evidence relative to hypothesis generation, firewalled data collection runs counter to the way that qualitative research is generally conducted. Instead of proceeding linearly from theory generation to data collection to theory testing, we naturally engage in a "dialogue with the data," (Bayesian astrophysicist Stephen Gull, quoted in Sivia (2006)) where we go back and forth between theory and evidence. We revise and refine theory in light of the data, and we revisit the evidence in light of new ideas and new theory, analyzing the information differently or more deeply, asking new questions, and deciding what kinds of additional data to collect.

Firewalled data collection would come at a significant cost of precluding an effective dialogue with the

---

1 As CGH (this issue) write: "The reference to 'veils of ignorance' arises from a division of labor that allows a research assistant to carry out key data selection and coding tasks without knowledge of the theories, hypotheses, and mechanisms being tested by the principal."