

Is causal-process observation an oxymoron? a comment
on Brady and Collier (eds.) *Rethinking Social Inquiry*

Nathaniel Beck*
Department of Politics
NYU
New York, NY USA
nathaniel.beck@nyu.edu

Draft of November 16, 2005

*Thanks to Henry Brady, David Collier and Gary King, as well as my colleagues Bernard Manin and Adam Przeworski for interesting conversations on this topic.

When I first read King, Keohane and Verba (1994), (“KKV”) I was excited by the general theme of what seemed to me the obvious, but usually unspoken, idea that all political scientists, both quantitative and qualitative, are scientists and governed by the same scientific standards. KKV, as Henry Brady (Brady and Collier, 2004, ch. 3) cleverly notes, clearly were preaching to the qualitative researcher, and the subject of the sermon was that qualitative researchers should adopt many of the ideas standard in quantitative research. Thus, to abuse both the metaphor and KKV, one might compare KKV to the view of the Inquisition that we are all God’s children. But, having sat through many talks based on case studies, it is clear to me that getting everyone to think about the nature of the scientific enterprise we are all engaged in can only be a good thing.

Brady and Collier (2004) (“BC”) clearly take no exception to this general position. But while KKV more or less argue that, *mutatis mutandis*, qualitative researchers should learn much about research design from quantitative scholars, BC put more emphasis on the *mutatis*, so that all God’s children can happily coexist, whether they worship one or many gods. My question for the current sermon is whether those many gods are sufficiently similar. In particular, I am concerned about the role of “causal-process observations” (“CPOs”) in the scientific endeavor.¹ The relationship of CPOs to more standard (from a quantitative perspective) “data set observations” (DSOs), is, in my view, the key innovation in BC.

BC (pps. 277–8) define a CPO as “[a]n insight or piece of data that provides information about context, process or mechanism, and that contributes distinctive leverage in causal inference. A causal-process observation sometimes resembles a “smoking gun” that confirms a causal inference in qualitative research, and is frequently viewed as an indispensable supplement to correlation-based inference in quantitative research as well.” The tie of CPOs to the qualitative analyst’s standard method is strengthened by adding to the definition a reference to “process tracing.” CPOs are distinguished from DSOs; the latter are the quantitative researcher’s typical measures on a set of variables for each “subject” or “case” in the study. The issue before us is not whether standard qualitative tools such as process tracing are of value for qualitative analysis, but whether they can be used to solve some problems in research design in a novel way.

KKV argue that many qualitative research designs suffer from having too few observations (“small N”) for which the remedy is to increase the number of observations (find a way to get a “large N”) or related problems.² While more DSOs would clearly solve the small N problem, this solution is typically not available to the qualitative researcher. But CBS (see

¹This topic appears in much of the original work in BC. For concreteness, this sermon is based primarily on the concluding chapter, and in particular pps. 252-66. Since the authors of the chapter are listed as Collier, Brady and Seawright, I will refer to it as “CBS.” BC contains much more than the discussion of CPOs. Thus an important sermon might be “what the inquisitor can learn from the heathen?” This would relate to the important and useful discussions in the book on the limitations of even the best quantitative methods which use observational data to make causal assessments, as well as mistakes quantitative researchers frequently make when they assume causal homogeneity.

²I will focus here on the CBS claim that CPOs allow the qualitative analyst to surmount the small N problem. BC contains related discussions on how CPOs also surmount related issues such as a lack of variation on some variable. To keep this sermon short, I only discuss issues related to N, though I would claim that this discussion holds equally for the other claims of BC about the merits of CPOs for solving research design issues.

footnote 1) argue that more DSOs are not the only solution to the problem; researchers can obtain “causal leverage” via gathering CPOs as well as DSOs.

Thus CBS (p. 261) consider a comparative analysis based on 24 observations. Obviously the researcher could increase causal leverage by adding more observations, but CBS note that this could cause more harm than good if the new observations violate “causal homogeneity.” They then note that an alternative would be to add to the analysis additional CPOs (on four critical observations), increasing causal leverage without changing N. Have CBS an alternative to KKV’s solution to the small N problem (‘increase N’)? To help answer this I looked at the three examples given by CBS. I start with the most qualitative example and end with the most quantitative one. I stress that all discussion of these works is based on the discussion in CBS, and it is CBS’s claims, rather than those of the original authors, that are at issue here.

The first example is a purely qualitative assessment of why the US has not used nuclear weapons after World War II (Tannenwald, 1999), an in-depth analysis of four cases. Obviously the N is tiny (and there is no variation on the dependent variable in the cases of interest). Increasing N is impossible. So how can the claim that nuclear weapons were not used because of a “nuclear taboo” be assessed? Tannenwald turned to documents about decisions to use nuclear weapons (process tracing). Does this allow for an inference about the use of nuclear weapons?

Qualitative researchers would have no trouble with Tannenwald’s evidence; her methodology is almost prototypical. The issue is how to make sense of combining this qualitative evidence with the (in this case meager) quantitative evidence. On this issue CBS are silent. As CBS note, Tannenwald’s evidence could be accounted for in other ways. They then conclude (p. 258) that “[h]owever, to the extent that researchers find alternative accounts such as strategic misrepresentation less plausible, Tannenwald’s causal-process observations provide valuable support for her argument.” To my mind this simply begs the question: good research design is about ruling out these alternative explanations, and CBS provide no account of how this is to be done in the context of this study. I have no reason to doubt that Tannenwald’s account of what decision makers put in memos is correct, that is, they clearly claimed, and probably thought, that the nuclear taboo was very important. But the question at issue is not “what did decision makers claim was important to them?” but rather “why did the US not use nuclear weapons?” These are two related, but different, questions.

The second study discussed is Stokes’ (2001) analysis of the determinants of economic policy in Latin America from 1982–1995. This is an excellent exemplar of large N comparative research (38 cases), supplemented by considerable qualitative discussion of the cases. Stokes concluded that presidents chose neo-liberal policies because they believed such policies would solve fundamental problems. So far we are in the realm of standard DSOs and no interesting issues beyond those considered in KKV have yet arisen.

CBS (p. 257) then note that Stokes “supplements this large-N analysis by examining a series of causal-process observations concerning three of the presidents....[H]er inferential leverage derives from the direct observation of causal links.” Thus she found that after his election, Fujimori encountered various leaders who exposed him to macroeconomic arguments which convinced him of the wisdom of adopting neoliberal policies. Clearly no one could be

opposed to knowing one's cases, and clearly all social scientists do (or should) look carefully at a variety of cases before and after more systematic quantitative analysis. But what do CBS mean by inferential leverage? Obviously we have all looked at some regressions and said they are nonsense, they simply miss some feature of the world; others seem more believable. Here one might trust Stokes since she clearly does understand her cases. But can we infer beyond what we could infer from the 38 case regression? I do not see how.

The Stokes discussion raises two points. The first is whether our interest is in finding some general lawlike statements or in explaining a particular event. Note that while Stokes' book is about why politicians change policies after being given an electoral mandate. But CBS (p. 257) go on to say that CPOs "thus provide valuable evidence for the argument that Fujimori's decision was driven by this [economic] conviction, rather than by rent-seeking concerns identified in the rival hypothesis." So the qualitative analysis is helpful for understanding one specific case, answering the question "why did Fujimori undertake the policies he did?" rather than the general question of why politicians change policies after being elected? For me it is the general question that is of interest, and so it is unclear how the qualitative analysis provides additional leverage (beyond the idea that one should know and understand one's cases).

The second relates to what it means to directly observe a causal link. We can ascertain whom Fujimori talked with, and we can certainly get accounts by those present as to why he did what he did. But should we take these at face value? While it is of interest to read the documents and hear the accounts of participants, it is difficult to know if these accounts are yielding the actual causal mechanism. It is not surprising that actors give accounts in terms of ideas, and that learning that policies are good policies is a common account of why policy-makers do what they do. As with Tannenwald, the issue is how we come to believe that these accounts are correct. Thus I do not know what it means to directly observe a causal process or how I know when a researcher observes such a process. Hence I think that Stokes' work can only be evaluated by the standards set forth in KKV. I do not believe that KKV ever suggested it was wrong for analysts to understand their cases, or for them to have done detailed studies to better understand those cases (or a subset of them).

The third study is Brady's reanalysis (which appears as an appendix to BC) of an unpublished study by [Lott \(2000\)](#) on the impact of the early call of Florida for Al Gore in 2000. Lott argued that the call, ten minutes before the polls closed in the Florida panhandle, cost George Bush 10,000 votes. Lott came to this conclusion based on a time-series-cross-section regression of turnout in Florida counties over 4 elections. This regression was about as atheoretical as a regression could be, using as independent variables only county and year dummies. This regression was then used to conclude that Republican turnout in the panhandle counties was 10,000 voters under what would have been expected. In terms of the regression, "what would have been expected" means "controlling for county and year" or "given average county turnout over several elections and overall Florida turnout over those elections."

What Brady takes issue with is not the 10,000 vote figure, but rather Lott's conclusion that the TV stations' early call was the cause. Brady argues that adding CPOs to the study would have led to a much more plausible estimate. The key feature here is that the call for

Gore came only 10 minutes before the polls closed, and that only people who would have voted right when the polls were closing could have been affected by the early call. Brady calculates this number in total as 4200 potential voters (Republicans and Democrats). Given that both Republicans and Democrats are in this figure, and prior quantitative standard quantitative work on media effects and turnout, Brady concludes that Bush lost fewer than 100 votes due to the early call. Based on the evidence presented in the Appendix, it appears to me that Brady clearly bested Lott in this debate. But our interest here is in the role of CPOs, not the debate between Brady and Lott.

As with the discussion of why Fujimori chose the policies he did, the debate between Brady and Lott is about one specific case. While Brady brings to bear general theoretical knowledge about media and voting (all based on standard quantitative methods), we are unlikely to even have enough quantitative information to know how whether people in the panhandle rushed would have rushed to the polls right as they closed, or how many actually heard the early call, or how many changed their minds about voting. Given that there is reason to be interested in this specific case, it clearly makes sense to use “softer” evidence, such as discussions with poll workers about whether people in the panhandle tend to vote right before the polls close. But, as social scientists, our interests should be on theories of turnout and media effect, theories which Brady brings to bear on the specific case.

It should be stressed that Brady does not take issue with Lott’s regressions; turnout in the panhandle in 2000 was lower than expected, at least when compared to overall Florida turnout and the past history of county turnout. On the basis of no evidence, Lott attributed this loss of voters to the early call; Brady makes what seems to be a more compelling argument that the 10,000 vote figure is a function of Democratic mobilization efforts in more Democratic leaning areas and amongst more likely Democratic voters. Obviously for the specific issue of what happened in 10 counties in Florida in 2000 we are unlikely to be able to use our standard national survey methods (exit polls obviously would not work, and the NES does not interview enough Panhandle voters). Perhaps some of the new Internet surveys which have a massive number of respondents might help settle the question, but such evidence either does not exist or has not been brought to bear on the question.

Brady concludes (p. 271) that “it would be hard to collect data that could rule out all the possible confounding effects. Consequently, rather than seeking additional data-set observations, in my judgement it would be more productive to do further in-depth analysis of causal-process observations drawn from these ten Florida panhandle counties, finding out what happened there, for example, by interviewing election officials and studying expert reports.” Interestingly, Brady fails to suggest what would have been closest to process tracing, that is, unstructured interview with non-voters in the panhandle. Note that we seldom study non-voters accounts of why they do not vote, perhaps because we think they would be self-serving. (Thus one could imagine an article, perhaps similar to Tannenwald’s, on accounts that non-voters give for their lack of voting, but it is hard to see what this would add to standard studies of turnout.)

Sometimes, perhaps for a court case, we need to make causal attributions for one specific case study. Some standard quantitative evidence (along with common sense, such as no one who voted before the call for Al Gore could have had their turnout affected by that call) can

be brought to bear on this specific issue. But, as in this case, it is likely that this standard quantitative evidence will be insufficient to settle the question for the specific case. While one could imagine an investigation of individual voting records in Florida (which might contain the time of day that people vote), it is unlikely that a researcher will, in practice, be able to do this. So depending on the accounts of election officials might be the best that can be done. But, as with the issue of Fujimori's policy choices, if we want to obtain knowledge about the relationship between variables of interest, it is simply hard to see how either CPOs could be adjoined to the data set, or how these CPOs can solve problems of a defective research design. Thus, if our question is, what is the effect of early calls for a candidate on turnout, and we have only two early calls to work with, we are simply not going to be able to answer the question, and no additional causal-process "observations" are going to make up for the fact that we have only two data-set observations (quotes and lack of quotes are intentional).

We can see this more clearly with a more standard example, one related to studying general relationships, not explaining one specific event. Suppose we regress Congressional vote for the incumbent on campaign spending by the incumbent. Suppose we find almost no relationship. We might conclude that money does not matter, and that everyone who thought that money did matter was wrong. This would be consistent with this regression. A Martian might choose to stop at this point.

But no student of elections would stop there. Theory would tell us that challenger spending matters, and perhaps increased incumbent spending is related to increased challenger spending. Or perhaps incumbents in trouble spend more to offset their troubles. The electoral analyst would then incorporate these theoretical ideas (and ideas which are also consistent with knowledge of the cases) into more appropriate regressions, which would then yield more believable results. The results would be more believable because they are based on models consistent with extant theory, and because they are consistent with our knowledge of the cases. Thus no one (other than our hypothetical Martian) relies *only* on DSOs; it is DSOs, in conjunction with theory and knowledge of the cases that allow us to obtain quantitative results that we believe. But to talk about CPOs, and to argue that one should pursue additional CPOs (as if they were like additional DSOs) simply does not get us very far. So we would surely hope that our electoral analyst has seen some real campaigns and talked with some real campaigners. How the information gleaned in such activities helps our regression runner is typically not dealt with in the standard econometrics texts, though it is a more central issue in the Bayesian world (Gill and Walker, 2005; Leamer, 1978). In any event, understanding what one studies is a good thing, but I doubt that such a statement would bother KKV.

Note that if we wanted to understand why some particular Incumbent X won without spending any money we might turn to a variety of explanations, some based on general laws or observed regularities, some based on idiosyncratic features. Perhaps we would talk with local experts or even journalists. But as social scientists our interest should be in generating and testing theoretical propositions, not explaining that incumbent X won without spending any money because challenger Y was caught *in flagrante delicto*. Much of what we actually do is what KKV call "descriptive inference." As we saw in the campaign spending example, theory can help us decide which regressions yield more valid descriptive inferences, but one

should not confuse this use of theory with making observations on a causal process, let alone joining such observations to an enhanced data set.

So my take is that BC have not really made it clear what it means to adjoin CPOs to DSOs. This undermines a variety of key points in the book, particularly as it relates to a more nuanced view of the research process and a much needed examination of the cleanliness of the hands of the quantitative analyst. Both KKV and BC agree that qualitative and quantitative researchers should share the same standards; BC make a noble effort to show that the tools of the qualitative analyst meet those standards; but I find that effort not quite as persuasive as do Brady and Collier.

REFERENCES

- Brady, Henry E. and David C. Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Md.: Rowman and Littlefield.
- Gill, Jeff and Lee D. Walker. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research." *Journal of Politics* 67:841–72.
- King, Gary M., Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Leamer, Edward. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.
- Lott, John. 2000. "Gore Might Lose a Second Round: Media Supressed the Bush Vote." *Philadelphia Inquirer*. November 14, p. 25A.
- Stokes, Susan Carol. 2001. *Mandates and Democracy: Neoliberalism by Surprise in Latin America*. New York: Cambride University Press.
- Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basis of of Nuclear Non-Use." *International Organizations* 53:433–68.