The Inference in Causal Inference:  
A Psychology for Social Science Methodology

John Gerring  
Boston University  
<jgerring@bu.edu>

Jason Seawright  
Northwestern University  
<j-seawright@northwestern.edu>

Alejandro Avenburg  
Boston University  
<avenburg@bu.edu>

Draft: June 14, 2012

Word count (without appendix): 6,300

Comments requested!  
Please do not cite without permission.
ABSTRACT

In this paper, we outline a new area of research dubbed the *psychology of methodology*. This field is concerned with the investigation of how inferences, and especially causal inferences, are reached by academics – in this case, members of the social science professions. We assume that a reader’s willingness to accept truth-claims made in a work of social science may be influenced by the nature of the evidence as well as by characteristics of the venue in which a study is published or presented, characteristics of the author, and characteristics of the reader.

As an exemplar of the sort of work that may be conducted in this area, we focus on a controversial element of contemporary social science – the distinction between qualitative and quantitative methods. Specifically, we ask how the nature of the evidence affects the sort of causal inferences that social scientists draw and the confidence with which they draw them. To do so, we subject a sample of (mostly doctoral) students to an experiment in which each species of evidence constitutes a distinct treatment, while all being oriented around a central argument. We then ask respondents a series of questions about that argument to gauge the inferences they are willing to draw. We find that the nature of the evidence has considerable impact on some aspects of causal inference, but not on others. In some respects, the pattern of differential responses conforms to widespread methodological norms about the strengths and weaknesses of qualitative and quantitative evidence. In other respects, it contradicts those norms, suggesting that social scientists may be systematically misinterpreting evidence when judging published work.
Many traditions of work converge on the topic of causation, each with a somewhat different focus. Philosophy focuses on the meaning of causation. Methodology focuses on how to achieve causal inferences, with the goal of maximizing validity and precision. History of science focuses on changes over time in methodological principles and practices. Sociology of science examines the contemporary practice of science, with an eye to social and political influences on the production of knowledge. Psychology focuses on how lay people – children or adults – understand the world in causal terms.

Our focus is nestled in between methodology and psychology. We want to know how social scientists reach causal inferences, as well as inferences of other sorts.¹ We shall refer to this field as the psychology of methodology.

It is probably apparent that social scientists and lay people use somewhat different tools to reach, and to justify, causal inferences. Indeed, most lay readers would be puzzled by journal articles in economics, political science, or sociology. The question of inference within the social science community thus provides a distinct topic of research. (This is not to deny that there is some overlap in causal reasoning between social sciences and lay people, which constitutes an interesting research agenda in its own right.)

Skeptics within the social science community might wonder whether research on perceptions of causality is helpful. After all, the point of social science is to reach valid and precise causal inferences, not to talk about the psychology of doing so. We want to know how to do it, and how to do it well, not how social scientists conceive of it. Worse, the label ‘psychology of methodology’ might seem to imply that approaches to causal inference are simply the product of cognitive or pre-cognitive biases, i.e., that there is no objective approach to causal inference.

Our goal is not to debunk the notion of scientific truth. To the contrary, we regard the psychology of methodology as an important aid to improving the practice of social science and the achievement of greater consensus within scientific fields. To the extent that social scientists’ actual, applied causal-inferential standards correspond with one or another position in methodological debates the stakes of those debates are clarified, as are the persuasive powers of various techniques. To the extent that psychological research on methodology reveals causal-inferential practices that do not correspond to articulated positions in methodology, that finding reveals clear opportunities for improving social-scientific practice.

As an example, we focus here on a controversial element of contemporary social science – the distinction between qualitative and quantitative methods. Specifically, we ask how the nature of the evidence – qualitative, quantitative, or multimethod – affects the sort of causal inferences that social scientists draw, and the confidence with which they draw them. To do so, we subject a sample of (mostly doctoral) students to an experiment in which each species of evidence constitutes a distinct treatment, while all being oriented around a central argument. We then ask respondents a series of questions about that argument to gauge the inferences they are willing to draw. We find that the nature of the evidence has considerable impact on some aspects of causal inference, but not on others. In some respects, the pattern of differential responses conforms to widespread methodological norms about the strengths and weaknesses of qualitative and quantitative evidence. In other respects, it directly contradicts those norms, suggesting that social scientists may be systematically misinterpreting evidence when judging published work.

We begin by describe the research question and main hypotheses. Next, we lay out the research design for this experiment. Finally, we present the findings. By way of conclusion, we review other areas of investigation that seem ripe for analysis, establishing a research trajectory for a

¹ A causal framework provides an overarching structure to our project, but it does not prevent attention to other kinds of inferences. Descriptive or predictive inferences often play a major role in, or rely on, causal arguments; as such, several of our arguments below apply to these kinds of inferences.
psychology of methodology.

I. Does Method Matter?
The qual/quant contrast has a long history – some might argue, too long. We realize that the concepts are difficult to define and the distinction perhaps overblown, providing a distraction from methodological issues that are more consequential (Gerring 2012; McLaughlin 1991). Nonetheless, “qual” and “quant” are ubiquitous terms in the contemporary social sciences and their very contentiousness suggests that the nature of the evidence may influence the way scholars evaluate an argument (Brady & Collier 2004; Glassner & Moreno 1989; Goertz & Mahoney 2012; Hammersley 1992; Rossman & Rallis 1998; Shweder 1996; Snow 1959/1993). As such, the qual/quant distinction constitutes an important locus of research in the psychology of methodology.

For present purposes, we define quantitative as work that employs large samples explored through inferential statistics. This genre is surveyed by textbooks on statistics and econometrics (Greene 2002), as well as more wide-ranging methodological texts that adopt a quantitative template for understanding research design (King, Keohane & Verba 1994). Qualitative, by contrast, will refer to empirical evidence that is small in numbers of comparable observations (though there may be a multitude of non-comparable observations), analyzed without use of formal statistical models. Data gathered through participant-observation, unstructured interviews, focus groups, or archival research generally has this quality – at least initially, prior to any subsequent coding that the researcher may choose to perform (Kapiszewski, MacLean & Read 2015). Likewise, research focused on a single case or a small number of cases generally leans heavily on qualitative evidence – though it may also incorporate quantitative evidence (Gerring 2007). Multimethod research will be understood as work that integrates both types of evidence in the same research product, i.e., article, report, or book (Seawright 2015).

Are there systematic differences in the functions and utility of qualitative and quantitative evidence? Methodological studies, reviewed below, suggest that there are. Among non-methodologists one may expect that methodological predilections follow methodological practices. That is, scholars whose work is primarily quantitative are most likely to prefer a quantitative approach to a subject; scholars whose work is primarily qualitative are most likely to prefer a qualitative approach; and scholars whose work embodies a multimethod approach are likely to prefer the latter. We shall assume that these preferences are not merely aesthetic; they should also translate into beliefs about the causal properties of a subject.

Of course, not all of the potential methodological implications of qual and quant are amenable to experimental tests. (For example, it might be difficult to test whether qualitative or quantitative data is more useful for exploratory research – where the goal is to generate new hypotheses about a subject or to identify a new research question.) And some of them pertain to descriptive inference. Here, we focus on methodological issues that are accessible to experimental analysis and that pertain to causal inference.

First, some methodologists appear to believe that one form of evidence is superior to the other for purposes of reaching causal inference for the cases under study, i.e., internal validity. Work that advocates the use qualitative methods often carries the implication that qualitative evidence is more trustworthy than quantitative evidence, even if the authors aren’t entirely explicit on the issue (Beach & Pedersen 2013). Work that is critical of qualitative methods, or advocates the expansion of sample size to solve inferential problems, conveys the opposite impression (Beck 2006, 2010; King, Keohane & Verba 1994; Lieberson 1985, 1992, 1994; Sekhon 2004). In sum, there is some reason to suppose that either qualitative or quantitative methods might provide greater internal validity, at least in some settings – though it is unclear in which direction the advantage lies.
Second, by virtue of its capacity to delve deeply into the circumstances of an individual case, or a small number of cases, qualitative evidence is generally regarded as providing greater insight into causal mechanisms – at least for the studied case(s). This does not mean that quantitative analysis is irrelevant; indeed, large-n analysis can often shed light on the pathways lying between X and Y. Yet, as a matter of practice, those features that lie “inside the box” are often impossible to measure systematically across a large sample, and hence can only be approached in a case-centered, qualitative fashion, i.e., process-tracing (Bennett & Checkel 2015; Gerring 2007; Waldner 2012).

Third, by virtue of its capacity to represent a larger population (in whatever respect is relevant for the argument at hand), quantitative evidence is often regarded as having greater generalizability (external validity relative to a broader population). This does not mean, of course, that large samples are always representative of a larger population; they may be heavily biased. However, the hallmark of external validity, random sampling, is generally more viable – and certainly more commonly practiced – in quantitative research. Likewise, stochastic error is minimized when samples are large (George & Bennett 2005; Lieberman 2005; Seawright & Gerring 2008).

Finally, a burgeoning literature on multi-method research suggests that in many settings qualitative and quantitative evidence perform complementary functions. If so, a multimethod approach to causal inference may combine the strengths of both sources of evidence, providing greater support for inference than either method employed on its own (Brannon 1992; Brewer & Hunter 2006; Clark & Creswell 2007; Malterud 2001; Morgan 2014; Seawright 2015; Teddlie & Tashakkori 2009; White 2002).

Thus, we arrive at four main hypotheses:

1. (a) Qualitative evidence is superior for attaining internal validity.
   (b) Quantitative evidence is superior for attaining internal validity.
2. Qualitative evidence is superior for understanding causal mechanisms.
3. Quantitative evidence is superior for attaining external validity.
4. Multimethod evidence is superior for all (or most) purposes.

II. An Experimental Design

An opportunity to test the hypothesis that scholars in fact evaluate real-world studies in line with the methodological guidance sketched above is provided by multimethod research where the quant and qual portions of the study are easily distinguished, i.e., occupying different sections of an article. In this context, it seems reasonable to inquire whether qualitative or quantitative evidence – or both combined – has greater impact on the reader’s overall assessments about purported causal relationships.

Our study begins with a selection of top empirically oriented journals in political science including American Journal of Political Science, American Political Science Review, British Journal of Political Science, International Organization, Studies in American Political Development, and World Politics. We exclude journals that are primarily methodological (e.g., Political Analysis) or theoretical (e.g., Political Theory). All articles in these journals across a decade (2002-2011, inclusive) are reviewed. Articles are included in an initial cull if all of the following criteria are satisfied:

1. At least half of the article is devoted to the presentation of empirical evidence. (Articles whose aim is primarily theoretical or methodological are excluded.)
2. The argument is causal and a principal hypothesis or set of closely related hypotheses can be identified. (Articles in which there are multiple, diverse hypotheses or which attempt to explain all factors contributing to an outcome, a causes-of-effects style of argument, are excluded.)
3. Both qualitative and quantitative evidence is offered in support of the hypothesis and both sections – qualitative and quantitative – are fairly clearly delineated (i.e., separated by headings).
4. The evidence is nonexperimental, which is to say that the treatment is not intentionally
randomized across groups.

5. The article is not written by the authors of this paper.

The resulting fifty-three articles are then classified into one of three categories: comparative politics ($n=28$), international relations ($n=19$), and American politics ($n=6$). From each category, five articles are chosen with a random number generator, with the caveat that no author should be represented more than once in the resulting sample. A list of these fifteen articles can be found in Table A1.

For each article, three treatments are then devised. The Multimethod treatment consists of the entire article, as published. The Quantitative treatment omits the qualitative section of the article and any discussion of it elsewhere in the article. The Qualitative treatment omits the quantitative section and any discussion of it elsewhere in the article. Tables with descriptive statistics and formal models are retained in all treatments. However, author’s names and any identifying features are removed. Other than these omissions, articles are presented exactly as they were published, using PDF’s downloaded from JSTOR or other on-line repositories. Omissions are created by blacking out a portion of the text (using Adobe Acrobat).

After reading a brief description of the objectives of the survey (which is entirely anonymous), respondents are asked to classify themselves into one of the three subfields. An article is chosen randomly from that subfield (from the pre-selected set of 5) and the respondent is presented with a brief description of the article’s main hypothesis about the causal effect and the causal mechanism, as well as a narrow context in which the argument might apply (addressing the issue of internal validity), and a larger context in which the argument might apply (addressing the issue of external validity or generalizability). For example, in Trejo’s (2009) study of religious organizations’ response to competition, the narrow context is specified as Mexico between 1975 and 2000 – effectively, the sample that Trejo studied – while the larger context is defined as all Latin American countries since the expansion of the Protestant Church in the region (1970s to the present).

No evidence is presented at this stage, whose purpose is to elucidate prior beliefs. Specifically, the respondent is asked whether s/he thinks the argument about the causal effect is true in the narrow context and in the broader context, and whether s/he thinks the argument about the causal mechanism is true in the narrow context and the broader context. After each answer, the respondent is asked about how confident s/he feels about his/her answer (on a scale of 0-100). This is the pre-test.

Next, the respondent is given one of three versions of the article to read – the Multimethod treatment, the Quantitative treatment, or the Qualitative treatment, as described above. The respondent is then asked whether s/he has previously read the paper, heard commentary about it, or knows the author(s) personally (and feels that they might be influenced by this fact). A “yes” response to this question triggers an automatic assignment to a different article. (In the event, no respondents answered in the affirmative, so no one was reassigned.)

After reading the article, the respondent is subjected to the same battery of questions about its argument. Do you think the argument about the causal effect is likely to hold in the narrow context (Y/N)? Do you think the argument about the causal effect is likely to hold in the broader context (Y/N)? Do you think the argument about the causal mechanism is likely to hold in the narrow context (Y/N)? Do you think the argument about the causal mechanism is likely to hold in the broader context (Y/N)? This is the post-test. A set of post-survey questions inquire about the respondent’s

---

2 www.random.org/

3 This means, of course, that respondents are aware of when and where a section of the text has been omitted. Overcoming this feature of the experiment would involve re-setting the type and considerable editing of the articles, so as to present the qualitative and quantitative treatments as stand-alone articles. Even so, we suspect that these reconstructed treatment conditions would lack the content and texture of published articles in top journals.
familiarity with the subject matter of the article, training, and demographic characteristics. The complete questionnaire is reproduced in Table A2.

The first iteration of the experiment was conducted at the annual Institute for Qualitative and Multi-method Research (IQMR), held at the Maxwell School of Syracuse University. The survey was administered in a plenary session of the Institute (including all attendees) as the first order of business on the first day of the Institute, June 18, 2012. The survey was timed at thirty-five minutes. One hundred and twenty-nine respondents completed essential portions of the survey. A second iteration of the experiment was conducted at IQMR on 15 June 2013, following the same protocol. Sixty-one respondents completed essential portions of the survey in this round. A third iteration of the experiment was conducted at the annual Summer School in Methods and Techniques sponsored by the European Consortium for Political Research (ECPR) at the University of Ljubljana, Slovenia, 27 July to 3 August, 2012. In this iteration, students were asked to take the survey on their own time (within the first week of the summer school). This means that students took the survey at different times (raising the possibility of contamination) and there is no time-limit on the duration of the survey. Twenty five respondents completed the survey in this round. Altogether, 215 students completed the experiment and filled in answers for the treatment and outcome variables of theoretical interest. This composes our sample.

Although they draw from different catchment areas, students at IQMR and ECPR are similar in most respects that might be expected to affect the results of this experiment. Almost all are enrolled in doctoral programs in the social sciences, among which political science predominates. A statistical portrait of this research group is contained in Table A3.

III. Findings and Interpretation

Statistical analyses summarized in Table 1 compare the three treatment conditions and various outcomes, organized in a fashion that highlights the most significant differences. All analyses utilize ordinary least squares regression with bootstrapped standard errors (2000 replications). Given the dichotomous nature of the outcome variable, some readers might expect to see a logit or a probit analysis, in place of the linear probability model. However, for analyses that include only the experimental treatment and the outcome using such a model would clearly be a step away from optimal research practices, since the difference in means (estimated by OLS regression) is uniquely justified by the research design (Freedman 2008). When adjusting for covariates, the linear-probability model provides a sensible estimate of the conditional expectation function without relying heavily on assumptions about the distribution of the error term to produce estimates, as do logit, probit, and other maximum-likelihood models. For these reasons, OLS regression arguably provides more direct and less assumption-laden causal inferences than a model designed specifically for limited dependent variables (Angrist & Pischke 2009: 94-107).

The first set of tests in Table 1 focus on a broad rendering of the argument, i.e., how it would apply to a large population of potential cases. We begin by looking at evaluations of the causal effect. Model 1 contrasts the Quantitative condition (the omitted category) with the two other treatment conditions. We find that respondents are more likely to believe that the claimed causal effect is true for the broadly defined population when the evidence is quantitative than when it is qualitative. The Multimethod condition is also negatively signed, with a slightly attenuated coefficient (relative to the Qualitative treatment). Quantitative evidence is apparently preferred to qualitative and multimethod evidence in this context, though the contrast with multimethod evidence is not quite as stark (and does not quite cross the threshold of statistical significance at 0.10).

The next set of tests focus on a binary treatment in which the Qualitative and Multimethod
conditions are coded as 0 and the Quantitative condition is coded as 1. Model 2 repeats the format of Model 1 with this binary treatment, which we regard as our benchmark model. The estimated coefficient suggests that respondents in this experiment are 14% more likely to believe that the relationship is causal when presented with quantitative evidence than when presented with qualitative or multimethod evidence.

Model 3 removes respondents from the sample who completed the survey very quickly (in less than 25 minutes). The estimated coefficient is considerably enhanced, as one might expect when those who are taking a casual attitude toward the exercise are removed from the sample.

Model 4 returns to the full sample, this time with pre-treatment variables measuring the session (IQMR1, IQMR2, or ECPR) and the article (dummies for each of the 15 articles) entered as covariates in the model.

The next set of tests focus on the causal mechanism at work in the claimed causal relationship – again with the broad scope of inference. Models 5-8 replicate the format of models 1-4. Here, we find that the superiority of the Quantitative treatment condition is even more marked relative to the Qualitative and Multimethod treatment conditions. Estimated coefficients from these models suggest that respondents are 20-29% more likely to believe that the putative causal mechanism is true for the broader population if they are presented with quantitative evidence (alone).

The final set of tests in Table 1 focus on narrowly defined scope-conditions. Here, we find similar results if we focus on the signs of the estimated coefficients. However, the differences across treatment conditions are not statistically significant at standard threshold levels. This holds whether respondents are asked about the purported causal effect (Models 9-10) or the causal mechanism (Models 11-12).
Table 1: Diverse Evidence as Stimuli to Causal Inference

<table>
<thead>
<tr>
<th>Scope</th>
<th>Causal feature</th>
<th>Sample</th>
<th>Effect</th>
<th>BROAD</th>
<th>Mechanism</th>
<th>NARROW</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full</td>
<td>Full</td>
<td>Slow</td>
<td>Full</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>10</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quant</td>
<td>[omitted]</td>
<td></td>
<td></td>
<td>[omitted]</td>
<td>[omitted]</td>
<td></td>
<td>[omitted]</td>
</tr>
<tr>
<td>Qual</td>
<td>-0.15*</td>
<td>-0.21***</td>
<td></td>
<td></td>
<td></td>
<td>-0.11</td>
<td>[omitted]</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Multimethod</td>
<td>-0.13***</td>
<td>-0.20**</td>
<td></td>
<td></td>
<td></td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td>(0.81)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Binary treatment</td>
<td>0.14* (0.07)</td>
<td>0.21** (0.09)</td>
<td>0.15* (0.08)</td>
<td>0.20*** (0.07)</td>
<td>0.29*** (0.09)</td>
<td>0.21*** (0.07)</td>
<td>0.07 (0.07)</td>
</tr>
<tr>
<td>(1=Quant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-treatment vars</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Obs</td>
<td>215</td>
<td>215</td>
<td>120</td>
<td>215</td>
<td>211</td>
<td>211</td>
<td>211</td>
</tr>
<tr>
<td>R²</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.10</td>
<td>0.04</td>
<td>0.04</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Ordinary least squares regression with bootstrapped standard errors, in parentheses, based on 2000 replications. *p<.10  **p<.05  ***p<.01 (two-tailed tests)  
Outcome: perceived probability of a causal effect or causal mechanism for a broadly defined or narrowly defined population.  
Sample: Full (all subjects that completed the survey for questions contained in the model), Slow (the Full sample excluding subjects who finished the survey in less than 25 minutes), Imputed (multiple datasets generated by the Amelia multiple-imputation algorithm).  
Pre-treatment variables: Session, Article.
The implications of these results vis-à-vis our initial hypotheses may be briefly summarized. We find no support for H1: neither qualitative nor quantitative evidence seems to enjoy an advantage with respect to internal validity. When an argument is couched in narrow terms – essentially, limited to the sample under study – respondents are indifferent to (or at least not strongly affected by) the style of evidence presented (see Models 9-12).

We find no support for H2: there is no difference across quantitative and qualitative treatments when the argument is framed in terms of causal effects versus causal mechanisms (compare Models 1-4 and 5-8 or Models 9-10 and 11-12). This runs counter to one of the essential canons of case study research, where it is widely believed that case studies – usually based largely on qualitative evidence – have greater value for elucidating causal mechanisms than for estimating causal effects.

We find some support for H3: quantitative analyses are more likely to be viewed as generalizable than qualitative analyses (see Models 1 and 5). This is plausible insofar as quantitative analyses often incorporate a larger tranche of evidence than qualitative analyses, and for survey designs often attempt to collect a representative sample from the population of interest. Yet most of the included articles in this study do not use surveys and instead rely on data drawn from convenience samples of countries or on narrow and perhaps unrepresentative time periods. In other words, the quantitative analyses presented here face effectively the same kinds of issues of generalization as the qualitative analyses. However, it may nonetheless be true that the standardization of units and observations required by quantitative analysis lends itself to greater external validity. In any case, in this respect scholars seem to draw heavily on conventional wisdom about the strengths and weaknesses of quantitative and qualitative evidence.

We find no support at all for H4: there seems to be no advantage in packaging qualitative and quantitative evidence together in a multimethod approach to causal inference. Indeed, multimethod research seems to be at a disadvantage in some settings. Note that quantitative evidence is more likely to be viewed as generalizable than the combination of quantitative and qualitative evidence. That is, adding a qualitative component to a quantitative analysis makes the study appear to have less external validity than the same study with only the quantitative analysis. (The estimated coefficients for the multimethod treatment in Models 1 and 6 are nearly identical to the estimated coefficients for the qualitative treatment.)

This result is all the more surprising given that our sample of respondents seems predisposed to favor a multimethod approach to social science. Sixty-four percent self-identify with the multimethod ideal (see Table A3) and 88% are drawn from a summer program in Syracuse entitled the Institute for Qualitative and Multimethod Research (IQMR). It seems likely that perceptions of non-generalizability might be even stronger among researchers whose predilections are not aligned with multimethod research.

One possible explanation for this curious result is that respondents exposed to the multimethod treatment are not paying attention to the quantitative component. Another possibility is that the inferential work normally performed by a quantitative analysis in allowing the argument to generalize is devalued by the addition of qualitative evidence, which perhaps tends to raise complications, caveats, and contingencies. In effect, the whole study comes to appear more parochial, more context-specific.

Whatever factor(s) is at work, we see no methodologically grounded rationale. The addition of qualitative evidence should not compromise a study’s external validity if all else is the same (as it is in our research design). Here, methodological rules (that we imagine most methodologists would agree upon) and norms-in-practice (as revealed by the experiment) seem to collide. And here, arguably, a psychology of methodology is uniquely informative, prompting us to think hard about the ways in which multimethod evidence is deployed and processed. Either multimethod researchers...
must do a better job of packaging their goods and/or end-users must do a better job of judging the product.

IV. A Prospective Field of Study

Having explored the relative impact of qualitative and quantitative evidence on the perceived truth-value of a study we now consider how the nascent field – which we have labeled the psychology of methodology – might be extended and developed.

In the natural sciences (including psychology), the psychology of science is a recognized – though extremely small – field of research.4 Studies have focused on the extent to which scholarly judgments are biased by the characteristics and beliefs of reviewers, e.g., confirmation bias (Mahoney 1977). Other work has looked at the degree to which judgments are affected by presumably irrelevant matters connected to a piece of research, e.g., the identity and university affiliation of the author (Peters & Ceci 1982), the quality of writing (Armstrong 1980), the language in which an article is written (Nylenna, Riis & Karlsson 1994), or whether the research is described as being “in press” (Mahoney et al. 1978).

In the social sciences, however, these sorts of questions have not been widely or systematically studied.5 We do not know whether biases discovered in the foregoing studies also obtain in anthropology, economics, political science, sociology, and allied fields. Extant studies suggest that there is a fairly low level of interrater agreement among journal reviewers in fields such as cultural anthropology and social psychology (Cicchetti 1991). But we do not understand the sources of that disagreement.

A number of additional issues remain largely unexplored in both natural science and social science settings. These concern the venue of a published work as well as characteristics associated with the author (e.g., status, demographic characteristics, and sex) and the reader (e.g., age, cohort, training, professional status, discipline, paradigm, specialized knowledge). We shall briefly discuss each of these areas of research as well as possible research designs that might be enlisted to test them.

Consider, first, the venue in which a paper appears. The impact of this central feature of the academic world may be tested experimentally by presenting a piece of research as (a) a conference paper, (b) a working paper, (c) a published paper in a second-tier journal, (d) a paper published in a top journal, or (e) a paper published in a top scientific journal such as Nature or Science. (To minimize potential confounders, the paper should be re-set in a standard typeface with no distinguishing characteristics that would suggest its origin.)

Consider, second, the status of the author. To test this thesis, an imaginary name might be chosen for an already published article. Experimental treatments may be devised by appending various descriptions of the author, e.g., as (a) a graduate student, (b) an assistant professor, or (c) a full professor. The gender, race, or national origin of an author may also influence responses to it. Here, one can alter the name of an author of an article – from one that is recognizably male to one that is recognizably female; from one that is recognizably white to one that is recognizably black or Hispanic; or from one that is recognizably native to one that is recognizably foreign.

Consider, finally, the characteristics of readers – the audience for academic work. Conceivably, the age of a reader affects his/her methodological proclivities. Perhaps older academics are more attuned to methodological subtleties that escape their younger colleagues. There may also

---

4 For reviews, see Mahoney (1979), Miller (2006). We include cognitive psychology and organizational behavior under the rubric of natural science.

5 A few nonexperimental studies exist, though these are generally fraught with problems of identification (e.g., Crane 1967).
be cohort effects. Note that methodological training has changed over time such that recent graduates may be receiving quite different educations from their forbears. It may be useful to classify respondents by year of completion of their (most recent) degree so as to gain leverage on this question. (So long as a sample contains a sufficient number of respondents whose age does not correspond with their cohort, these two factors can be differentiated empirically.) Inferences may also be affected by the degree of training received. Here, one might compare responses by those with different levels of educational attainment – BA, MA, or PhD. Or, one might count the number of methodologically-focused classes an individual has completed. Inferences may be affected by one’s current professional status, e.g., (a) students, (b) recent graduates, (c) faculty, (d) researchers, or (e) working in some other (nongraduate) sector. Inferences may be affected by the nature of one’s discipline or sub-discipline. Inferences may be affected by the methodological/epistemological paradigms that social scientists identify themselves with. Inferences may be affected, finally, by the specialized knowledge that social scientists possess about the context of a particular piece of research. Local knowledge may matter. (Perhaps respondents with greater knowledge of the research area in which the article falls will show lower variance across responses.) These sorts of questions are not amenable to experimental manipulation; nonetheless, questions about reader background are not especially prone to confounders.

To our knowledge, none of these issues have been studied in a systematic fashion in the social science disciplines. Yet, they are eminently tractable, as our short discussion illustrates, and presumably informative. Knowing the answers to these questions would help us understand the degree to which scholarly judgments are affected by non-scholarly considerations, and the sort of biases to expect. They may also point the way to measures that might overcome, or at least minimize, these biases in our collective judgments.

---

6 A number of questions may be asked of respondents in order to gauge their degree of familiarity with the substantive (not methodological) context of a piece of research. One may construct a Likert scale in which respondents rate their own knowledge of an area. One may ask a series of factual questions about that area. Or one may ask whether respondents are conversant with a series of key published works in that area.
V. References


Mahoney, James, Gary Goertz. 2006. “A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research.” Political Analysis 14:3 (Summer) 227-49.


Appendix A
Table A1: Articles, Subjects, Treatments, and Scope-Conditions

<table>
<thead>
<tr>
<th>Article</th>
<th>Subjects</th>
<th>Treatments</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>American Politics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondents: 8.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument (causal effect):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Where offshoring (the relocation of the provision of a service or part of a manufacturing process to another country) has differential effects on high- and low-skill workers in an industry this often leads to divergent political demands among workers with different skill levels, creating incentives to organize and lobby separately.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Multimethod: 1421 lines.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quantitative: 1341 lines (94.37%). Quantitative analysis of labor groups’ decisions on whether to support FTAC’s (Film and Television Action Committee) initiatives to stop offshoring in the film industry since 1998 (144 cases).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Qualitative: 958 lines (67.41%). Narrative account on the history of FTAC and its relations with labor groups in the film industry since 1998 based on newspaper articles, FTAC’s documents and secondary sources (1 case)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Narrow: Hollywood, California, since 1998.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Broad: All countries with a service industry where offshoring is a significant phenomenon.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondents: 6.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument (causal effect):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Indian (Native American) tribes with high levels of government institutionalization or economic development succeed more often in their interactions with county governments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Multimethod: 2047 lines.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quantitative: 1730 lines (84.51%). Quantitative analysis of interactions between Native American tribes and U.S. counties (12 Native American tribes) from 1990 to 2000.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Qualitative: 1528 lines (74.64%). Narrative account of interactions using minutes of the meetings between Native American tribes and U.S. counties (12 Native American tribes) from 1990 to 2000.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Narrow: American Indian governments in the Southwest, Pacific Northwest, and Upper Plains from 1990 to 2000.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Broad: All interest groups who attempt to influence political outcomes at the local level in the United States.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondents: 2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument (causal effect):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• The interruption of trade between the U.S. and the Allied powers caused U.S.’ intervention in World War I.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Treatments:
- **Multimethod:** 1348 lines.
- **Quantitative:** 1157 lines (85.83%). Quantitative analysis of US congressmen’s roll-call votes in four initiatives affecting U.S. involvement in World War I in 1916 and 1917 (1 case).
- **Qualitative:** 771 lines (57.19%). Analysis of impact of World War I on the U.S. economy (1910-1918) based on official statistics and on secondary sources (1 case).

**Scope:**
- **Narrow:** The U.S. during World War I.
- **Broad:** All countries with significant military power and economic interests affected by war in modern times.


**Respondents:** 7.

**Argument (causal effect):**
- The G.I. Bill's educational provisions for veterans increased rates of political participation among these recipients.

Treatments:
- **Multimethod:** 1691 lines.
- **Quantitative:** 1492 lines (88.23%). Quantitative analysis of a survey of World War II veterans on their political participation after the war (716 cases).
- **Qualitative:** 1095 lines (64.75%). Semi-structured interviews with 28 World War II Veterans.

**Scope:**
- **Narrow:** Beneficiaries of the G.I. Bill in the U.S. after World War II.
- **Broad:** Beneficiaries of all public policies that promote social rights in the U.S.


**Respondents:** 5.

**Argument (causal effect):**
- Higher levels of religiosity cause lower levels of support for social insurance.

Treatments:
- **Multimethod:** 3030 lines.
- **Quantitative:** 2750 lines (90.76%). Quantitative analysis of the adaption of workers’ compensation programs in the U.S. states (48 cases) and analysis of two surveys combined (6,197 individuals), 1910-1939.
- **Qualitative:** 1374 lines (45.35%). Narrative account on the positions of the Catholic and Protestant churches towards social security (2 cases), 1910-1939.

**Scope:**
- **Narrow:** Workers’ compensation legislation enacted by state governments between 1910 and 1930 and New Deal unemployment relief.
- **Broad:** All countries in the contemporary period.

**Comparative Politics**


**Respondents:** 30.

**Argument (causal effect):**
- Legislative seat volatility (the total change in the percentage of seats or votes won or lost by all parties between elections) is highest in countries where either no social group is large enough to form a majority on its own, or a majority group contains within itself a second smaller majority group; it is lowest in countries where one, and only one, group forms a majority.

**Treatments:**
- *Multimethod*: 981 lines.
- *Quantitative*: 717 lines (73.09%). Quantitative analysis of legislative seat volatility in 33 African countries in the post 1990s.
- *Qualitative*: 678 lines (69.11%). Analysis of three African countries in the post 1990s using secondary literature.

**Scope:**
- *Narrow*: Sub-Saharan African countries in the post-independence era.
- *Broad*: All countries where ethnic identity is politically salient.


**Respondents:** 31.

**Argument (causal effect):**
- Proportional, multiparty systems and federalism encourage a "corporatist" (consolidated, highly centralized) form of organization among employer associations. By contrast, nonproportional, two-party systems and unitary constitutions encourage a "pluralist" (fragmented, decentralized) form of organization among employer associations.

**Treatments:**
- *Multimethod*: 2104 lines.
- *Quantitative*: 1503 lines (71.44%). Quantitative analysis of data from 16 countries from 1900 to 1930s.
- *Qualitative*: 1546 lines (73.48%). Case histories of employers associations in two countries in the late nineteenth century and early twentieth century based on original organizations’ documents, contemporary press and secondary sources.

**Scope:**
- *Narrow*: Advanced industrial countries during the early decades of the Twentieth Century.
- *Broad*: All democratic nations during the Twentieth Century.


**Respondents:** 29.

**Argument (causal effect):**
- Successful nation building, involving the integration of diverse ethnic groups into a single national identity (which may coexist with ethnic identities), leads to greater provision of public goods.

**Treatments:**
- *Multimethod*: 1494 lines.
- *Quantitative*: 1430 lines (95.72%). Quantitative analysis of district, village, household and individual level indicators in two districts of two African countries in the post 1990s.
- *Qualitative*: 642 lines (42.97%). Analysis of structured interviews in two African countries in the early 2000s.

**Scope:**
- *Narrow*: Kenya and Tanzania after independence.
- *Broad*: All developing countries with strong ethnic identities.

**Respondents:** 22.

**Argument (causal effect):**
- Economic dependence on oil disempowers women in the political sphere.

**Treatments:**
- **Multimethod:** 1733 lines.
- **Quantitative:** 1572 lines (90.71%). Quantitative analysis of oil production and employment data of 169 countries from 1960 to 2002.
- **Qualitative:** 1114 lines (64.28%). Case studies of 3 countries (one oil rich and two oil poor) since the 1950s using secondary sources.

**Scope:**
- **Narrow:** Algeria, Morocco, and Tunisia in the post World War II era.
- **Broad:** All countries in the post Second World War era.


**Respondents:** 33.

**Argument (causal effect):**
- In an impoverished multiethnic region monopolized by a single religious provider, the entrance of a new religious provider spurs the former monopoly provider to reach out to neglected constituencies (e.g., the poor and ethnic minorities) and help them mobilize politically.

**Treatments:**
- **Multimethod:** 2224 lines.
- **Quantitative:** 1608 lines (72.30%). Quantitative analysis of protests and religious competition in 883 municipalities in Mexico between 1975 and 2000.
- **Qualitative:** 1361 lines (61.20%). Life histories of three religious leaders and their interaction with the local communities in Mexico during the second half of the twentieth century based on secondary literature.

**Scope:**
- **Narrow:** Mexico between 1975 and 2000.
- **Broad:** All Latin American countries since the expansion of the Protestant Church in the region (since the 1970s).

### International Relations


**Respondents:** 12.

**Argument (causal effect):**
- Democratization causes an increase in public education spending, a reduction in private education spending, and a focus on universal primary education rather than elite-targeted tertiary education.

**Treatments:**
- **Multimethod:** 1465 lines.
- **Quantitative:** 1148 lines (78.36%). Quantitative analysis of political and educational data from 113 countries from 1960 to 2000.
- **Qualitative:** 973 lines (66.42%). Case studies of three countries in the second half of the twentieth century using secondary literature and statistical indicators.

**Scope:**

**Respondents:** 18.

**Argument (causal effect):**
- In seeking to modify repressive behavior by states, (a) preferential trade agreements with hard standards that tie material benefits of integration to compliance with human rights principles are generally effective, while (b) softer human rights agreements and (c) preferential trade agreements with soft human rights standards that are not tied to market benefits are not generally effective.

**Treatments:**
- **Multimethod**: 1584 lines.
- **Qualitative**: 1212 lines (76.52%). Narrative account of how preferential trade agreements influenced state behavior through coercion in 8 cases since the 1990s using official documents and secondary sources.

**Scope:**
- **Narrow**: All states in the last quarter of the 20th century.
- **Broad**: All states since the end of World War II.


**Respondents:** 22.

**Argument (causal effect):**
- Socialization-based methods (“persuasion”) applied by international institutions rarely change state behavior, especially when there is significant domestic opposition to a policy change. By contrast, membership conditionality methods imposed by international institutions (where an international organization such as the EU requires a change of policy as a condition of membership) often alter state behavior.

**Treatments:**
- **Multimethod**: 1401 lines.
- **Quantitative**: 804 lines (57.39%). Quantitative analysis of influence of various international organizations in domestic minority-related legislation in 4 European counties in the 1990s.
- **Qualitative**: 1229 lines (87.72%). Qualitative analysis of influence of these international organizations in domestic minority-related legislation in 4 European countries in the 1990s using elite interviews, official documents, correspondence among leaders and media reports.

**Scope:**
- **Narrow**: Policies adopted by the Organization for Security and Cooperation in Europe (OSCE), the European Union (EU), and the Council of Europe (CE) towards Latvia, Estonia, Romania, and Slovakia during the Nineties.
- **Broad**: The policies of any international organization (e.g. NATO, WTO, OAS, Mercosur) towards any country in the postwar era.


**Respondents:** 27.

**Argument (causal effect):**
Democratization results in higher risk of war in a partial transition (i.e., where a regime moves from autocracy to semi-democracy). This may be contrasted with a state of continuing autocracy, continuing democracy, transition to full democracy, or transition towards autocracy.

Treatments:
- Multimethod: 1845 lines.
- Quantitative: 1707 lines (92.52%). Quantitative analysis using indicators of political regime and participation in war of 243 countries during the nineteenth and twentieth century.
- Qualitative: 663 lines (35.93%). Narrative accounts of two cases of war among democratizing countries in the late nineteenth century and late twentieth century using secondary sources.

Scope:
- Broad: All states since the French Revolution.


Respondents: 15.

Argument (causal effect):
- In situations of civil war, refugee flows from the affected country cause an increase in the risk of subsequent conflict in both the host and origin countries.

Treatments:
- Multimethod: 1305 lines.
- Quantitative: 1228 lines (94.10%). Quantitative analysis using indicators of refugee flows and armed conflict in 174 countries since 1951.
- Qualitative: 813 lines (62.30%). Narrative account of 3 cases of refugee flows affecting stability since the 1970s using nongovernmental organization’s documents, press reports and secondary sources.

Scope:
- Narrow: All states in the postwar era.
- Broad: All states in the 19th and 20th century.
**Table A2: Questionnaire**

**Page 1: Introduction and Consent**

The purpose of this study is to better understand the impact of social science research on social scientists’ perceptions of truth.

The survey should take about 30 minutes.

It is anonymous and confidential. At no time will you be asked to reveal your identity.

Participation is strictly voluntary. If you don’t wish to participate you may exit from the survey at any time.

By participating in this survey, you will be given the opportunity to win a drawing for a $300 US gift card. To enter the drawing, complete the survey and provide your email address at the end. That email address will be used only for purposes of the drawing, and will be deleted from our databank as soon as the prize has been awarded.

If you have any questions you may contact [name omitted].

You may obtain further information about your rights as a research subject by calling the Boston University CRC IRB Office at 617-358-6115.

**Page 2: Protocol**

The protocol of the survey is as follows.

1. We ask about your research area (subfield) of interest.
2. We present you with some arguments drawn from a recently published paper (from your chosen subfield) in a top political science journal.
3. We ask whether you think these arguments are true.
4. We ask you to read the article itself – or some portion of it.
5. We ask again whether you think these arguments are true.
6. We ask a short battery of questions about your training, personal background, etc.

Needless to say, there are no correct or incorrect answers. This is not a test of intelligence or of your methodological or substantive knowledge. It is simply a test of how much your views of a topic are impacted by reading a social science article on the topic.

Please note that the only alteration we make in the original published article is to remove the names of the author(s) and – in some cases – a portion of that article. If a portion is removed the text will be blacked out, so it will be apparent to you where this removal has occurred.

In any case, please treat the article, as it is presented, as the only relevant information. There is no need to speculate on the identity of the authors or on what might have been contained in the removed section. Answers to questions in Step 5 should reflect only the information that you have obtained from reading the article (along with whatever background knowledge you might have about the context, the theory, and the method).
Page 3: Subfield filter

What is your primary area of research and training?

(a) Politics between states (“international relations”)
(b) Somewhere other than the United States (“comparative politics”)
(c) The United States (“American politics”)

Subjects are randomly assigned to three groups, blocking by their answers to the previous question. Practically, this means that the first (a) is placed in group I, the second (a) in group II, and so forth – so that balance across subfields is achieved across groups.

Page 4: Pre-test

Here is an argument about a causal effect drawn from a recently published article:

X causes(ed) Y.

Here is a possible causal mechanism(s) that might explain the relationship:

X causes(ed) Y because…

Here is a narrow context in which the argument might hold (i.e., cases which the causal factor, if administered, would have - or would have had - the hypothesized effect):

Nation-states within Latin America in the postwar era.

Here is a broader context in which the argument might hold (i.e., cases which the causal factor, if administered, would have - or would have had - the hypothesized effect):

All nation-states in the postwar era.

Do you think the argument about the causal effect is likely to hold in the narrow context?

(a) Yes
(b) No

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Do you think the argument about the causal effect is likely to hold in the broader context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Do you think the argument about the causal mechanism is likely to hold in the narrow context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Do you think the argument about the causal mechanism is likely to hold in the broader context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Page 5: Filter for prior knowledge of the paper
Below you will find the redacted abstract of a paper that develops the argument we discussed on the last page. The authors’ names have been removed, as has any identifying information.

To the best of your knowledge, have you previously read this paper, heard commentary about it, or know the author(s) personally (and feel that your response might be influenced by this fact)?

(a) No
(b) Yes

[If Yes is chosen, another article from the chosen subfield will be randomly assigned and the respondent will begin again at page 4.]

Page 6: Exposure to stimulus and post-test

The text of a recently published paper is included below. After reading the paper carefully, we would like you to once again answer the questions we asked earlier about the strength of the argument. These questions are available below the text.

Versions…
(a) Control: unabridged.
(b) Treatment 1 (qualitative): quantitative section removed.
(c) Treatment 2 (quantitative): qualitative section removed.

Here, again, is the proposed causal effect:
\[ X \text{ causes(ed) } Y. \]

Here is the proposed causal mechanism(s) that might explain the relationship:
\[ X \text{ causes(ed) } Y \text{ because…} \]

Here is a narrow context in which the argument might hold (i.e., cases which the causal factor, if administered, would have - or would have had - the hypothesized effect):
\[ \text{Nation-states within Latin America in the postwar era.} \]

Here is a broader context in which the argument might hold (i.e., cases which the causal factor, if administered, would have - or would have had - the hypothesized effect):
\[ \text{All nation-states in the postwar era.} \]

Do you think the argument about the causal effect is likely to hold in the narrow context?
(a) Yes
(b) No

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Do you think the argument about the causal effect is likely to hold in the broader context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Do you think the argument about the causal mechanism is likely to hold in the narrow context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

25
Do you think the argument about the causal mechanism is likely to hold in the broader context?

How confident do you feel about this answer? (Enter a number from 0 to 100, where 0 represents zero confidence – it is as likely to be true as false – and 100 represents absolute certainty.)

Page 7: General questions

Here are some general questions pertaining to the article that you read.

Did you have sufficient time to read the article carefully?
   (a) No, I could have used more time
   (b) Yes, I had plenty of time

How familiar are you with the research topic of the article you read?
   (a) Not at all familiar
   (b) Somewhat familiar
   (c) Very familiar

Did you have trouble following the arguments and the evidence as presented in the article (or portions thereof) that you read? [Select all that apply]
   (a) The presentation was hard to follow
   (b) The theory did not make sense to me
   (c) The methods were hard to understand
   (d) I understood the arguments and the evidence for the most part
   (e) I fully understood the arguments and the evidence

How would you evaluate the research design (including methods of analysis) presented in the article (or portions thereof) that you read?
   (a) The research design was not clearly laid out
   (b) The research design was weak, raising serious threats to causal inference.
   (c) The research design was weak but probably the best that could be devised under the circumstances.
   (d) The research design was quite strong, though not above reproach.
   (e) The research design was very strong.

If qualitative methods were employed in the article that you read, do you think these methods were well-suited to the research question?
   (a) No.
   (b) For the most part.
   (c) Yes.

If quantitative methods were employed in the article that you read, do you think these methods were well-suited to the research question?
   (a) No.
   (b) For the most part.
   (c) Yes.

If you read a version of the article that had parts (other than author names) blacked out, please answer the following question: Did you find that the article made sense without the omitted section?
   (a) No, the truncated article did not make much sense. For that reason, it was difficult to answer the
(b) Yes, the truncated article made sense. Even though parts were missing it could be evaluated on its merits. Thus, it was possible to answer the questions asked in this survey.

Page 8: Personal background

Here are some questions pertaining to your personal background.

In what year were you born?
[drop-down menu]

What is your sex?
(a) Male
(b) Female

What is the highest degree you have completed?
(a) High school/secondary school
(b) Undergraduate university degree (BA, BS, or equivalent)
(c) Professional post-graduate degree (Masters, MPhil, or equivalent)
(d) Doctorate (PhD or equivalent)

Are you currently pursuing a degree?
(a) No
(b) Yes

If yes, what degree is it?
(a) Undergraduate university degree (BA, BS, or equivalent)
(b) Professional post-graduate degree (Masters, MPhil, or equivalent)
(c) Doctorate (PhD or equivalent)

In what year did you enter your current (or highest) degree program
[drop-down menu]

What is the discipline of your current or highest degree program?
(a) Anthropology
(b) Education
(c) Economics/Business
(d) Political science
(e) Public policy/Public administration
(f) Social work
(g) Sociology
(h) Other

What is the name of the university at which you received, or will receive, your current (or highest) degree?
[open text]

How many courses have you completed at any level (undergraduate or graduate) that focus on some aspect of social science methodology (statistics, research design,...)? (Do not include workshops.)
[numeric]

How many courses have you completed at any level (undergraduate or graduate) dealing primarily with
Which of the following traditions best describes your view of social science methodology? [Choose all that apply.]

(a) Post-positivist, post-structuralist (e.g., Michel Foucault, Jacques Derrida)
(b) Interpretivist (e.g., Clifford Geertz)
(c) Positivist (e.g., King, Keohane, and Verba’s Designing Social Inquiry)
(d) Qualitative or multimethod (e.g., Henry Brady and David Collier’s Rethinking Social Inquiry)
(e) Qualitative Comparative Analysis (QCA) (e.g., Charles Ragin)
(f) Experimental (e.g., Donald Campbell)
(g) None of the above

If you have any suggestions for how we could improve this survey, and its general approach to understanding social science inquiry, please let us know.
[open field]
**Table A3: The Research Group**

Subfield:
(a) Comparative politics  53%
(b) International relations  35
(c) American politics  12

Familiarity with the topic addressed by the article:
(a) Not at all familiar  32%
(b) Somewhat familiar  55
(c) Very familiar  13

Age (mean):  31

Sex:
(a) Male  47%
(b) Female  53

Highest degree completed:
(a) High school/secondary school  1%
(b) Undergraduate university degree (BA, BS, or equivalent)  4
(c) Professional post-graduate degree (Masters, MPhil, or equivalent)  86
(d) Doctorate (PhD or equivalent)  7

Currently pursuing a degree:
(a) No  8%
(b) Yes  92

Discipline:
(a) Anthropology  0%
(b) Education  0
(c) Economics/Business  2
(d) Political science  80
(e) Public policy/Public administration  8
(f) Social work  0
(g) Sociology  5
(h) Other  5

Courses taken in some aspect of social science methodology (mean):  5

Courses taken in statistics or quantitative research methods (mean):  4

Methodological tradition (multiple answers possible):
(a) *Post-positivist, post-structuralist* (e.g., Michel Foucault, Jacques Derrida)  13%
(b) *Interpretivist* (e.g., Clifford Geertz)  20
(c) *Positivist* (e.g., King, Keohane, and Verba’s *Designing Social Inquiry*)  31
(d) *Qualitative or multimethod* (e.g., Brady/Collier’s *Rethinking Social Inquiry*)  64
(e) *Qualitative Comparative Analysis (QCA)* (e.g., Charles Ragin)  23
(f) *Experimental* (e.g., Donald Campbell)  12
(g) *None of the above*  7
**Table A4:** Descriptive Statistics for Outcome Variables

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$=0$</td>
<td>$=1$</td>
<td>Total</td>
</tr>
<tr>
<td>Broader causal effect</td>
<td>105</td>
<td>110</td>
<td>215</td>
</tr>
<tr>
<td>Narrow causal effect</td>
<td>160</td>
<td>54</td>
<td>214</td>
</tr>
<tr>
<td>Broader mechanism</td>
<td>91</td>
<td>121</td>
<td>212</td>
</tr>
<tr>
<td>Narrow mechanism</td>
<td>142</td>
<td>66</td>
<td>208</td>
</tr>
</tbody>
</table>