Transparency and Reproducibility in Multi-Method Research

Transparencia y replicabilidad en investigaciones con métodos múltiples

Dossier

Ciencia
Política

THAD DUNNING

University of California, Berkeley

FERNANDO ROSENBLATT

Universidad Diego Portales[*]

I. INTRODUCTION

Recent years have witnessed at least two major trends in empirical social-scientific research. The first is a heightened emphasis on experimental methods, the use of which has grown markedly in political science generally (Figure 1) as well as the study of Latin American politics in particular. Research designs related to the experimental approach are also increasingly prevalent: for example, nearly 100 articles using one type of natural experiment (the regression-discontinuity design) have been published annually in peer-reviewed economics and political science journals since the mid-2000s (Bueno et al. 2014).

Partially in connection with this "experimental turn," however—but partly independent of it—disciplinary discussion has also focused on a second development: new procedures and practices that seek to boost the transparency and reproducibility of empirical research. As we suggest in this introduction, and as the other contributions to this special issue of *Revista de Ciencia Política* (*RCP*) suggest, experiments can in principle improve the simplicity and credibility of empirical research. Yet in practice a host of ancillary methods, standards, and practices are attain those objectives. The focus on research transparency and reproducibility has therefore emerged not only as a corollary of the prominence of experiments, but perhaps also as a reflection of their limitations. Thus, concerns about publication bias, private data, and un-reproducible results have become much more prominent. Social scientists have focused increasing attention on possible remedies, such as study pre-registration, data sharing, and results-blind review.

Figure 1: Experimental Articles in APSR
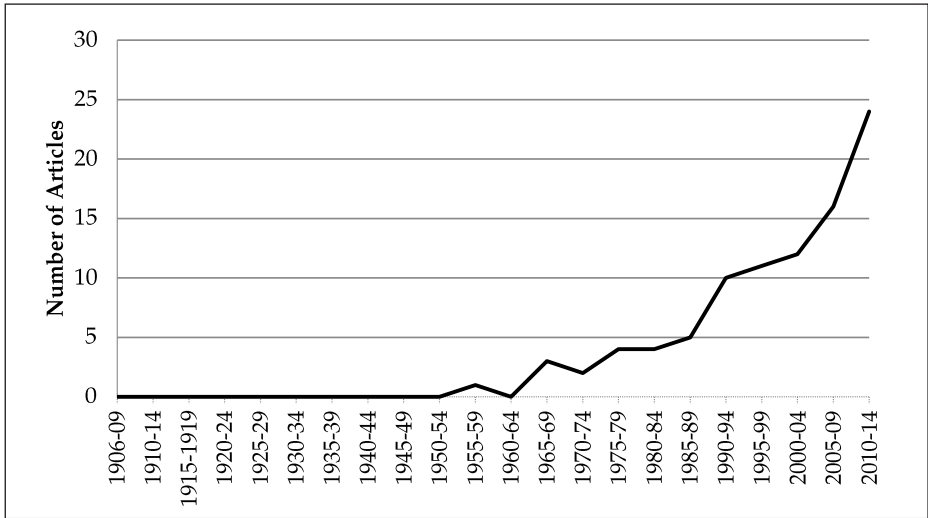
Source: Until 2005, Druckman et al. (2006); thereafter, JSTOR search using Druckman's criteria: "Requires primary data from a random assignment study with participants."

These two trends are welcome from several perspectives. As one tool in a multi-method toolkit, experiments and related designs can contribute to boosting the credibility of causal inferences in empirical research. The use of experiments—and "design-based research" more generally—has also led many political scientists away from reliance on complicated quantitative models towards simpler, more credible analysis. Meanwhile, the movement towards greater transparency can contribute substantially to our capacity for cumulative learning from distinct bodies of research.

Yet, these trends also leave many questions unanswered. Emerging practices such as pre-registration and data sharing may appear to have the greatest affinity with experimental research designs. Experimental protocols are developed in advance of interventions and data collection, making study registration feasible and natural. Yet, what are the implications of study pre-registration for research traditions for which such deductive pre-planning may appear a less obvious fit—including, very centrally, qualitative research? How can public commitments to research protocols be reconciled with the induction that is so critical for learning? In what ways can qualitative and quantitative methods be used together to improve the quality of experimental research designs? And what concrete tools and practices foster greater reproducibility? Social scientists have recently been vigorously debating such questions in a variety of fora, indicating their vital importance for research practice.[2] Answering such

---

[2]   For discussions on the trade-offs between research transparency, private data and un-reproducible results in qualitative research see Elman, Kapiszewski, and Vinuela (2010), Moravcsik (2014) or Büthe et al. (2015). See also the Data Access & Research Transparency statement at DART (http://www.dartsatement.org).

questions is critical not only for social science in general but also for the study of Latin American politics—which has an especially important tradition of qualitative and mixed-method research—in particular.

This special issue of *RCP* makes several contributions to this debate. In this introductory article, we briefly review the recent movement towards stronger research design and greater research transparency, focusing both on the goals these trends advance and the challenges they face. We emphasize especially the importance of multi-method research for building strong designs: the fusion of qualitative and quantitative methods can together contribute to improving not only experimental research but also non-randomized observational studies. This is a theme Maldonado et al. also develop in their contribution to this special issue. In their study of the reconstruction of the Chilean town of Chaitén in the wake of a devastating volcanic eruption, these authors demonstrate the utility of qualitative methods for deepening understanding the "selection" process that assigns units to "treatment" and "control" groups, showing how qualitative interview methods and quantitative matching analyses can complement each other for stronger causal inference.

Next, we also discuss the possible utility of public commitments to research protocols in qualitative research. This is a topic that Piñeiro and Rosenblatt deepen in their article on what they call Pre-Analysis Plans-Qualitative (PAP-Q). As Piñeiro and Rosenblatt argue, pre-analysis plans can play an important role in describing initial theory and usefully outlining process-tracing tests, thereby improving the quality and credibility of empirical research. Critically, however, study pre-registration must be designed in a way that allows for inductive learning as research questions and hypotheses evolve. Piñeiro and Rosenblatt indicate the usefulness of rolling amendments to pre-analysis plans, a practice that some experimental researchers have also adopted. These practices can improve the transparency of empirical research but also leave open some important questions for exploration—such as the feasibility of results-blind review for qualitative studies.

Finally, we consider practices that boost the reproducibility of research. Many scholars and institutions have long called for greater data sharing, a core element of open science; while challenges remain, there is some evidence of recent progress in political science. However, data sharing can involve a great range of practices, with greater or lesser specificity in terms of the specific codes and procedures that are used in a given analysis. It is typically not sufficient to simply distribute a data set without supplementary explanation that is useful not only to the broader community of scholars but also often to the original authors themselves. Bowers and Voors illustrate some of the challenges and share specific practices that can help our "future selves"—as well as other scholars—better understand and therefore reproduce specific design and

analysis decisions. These very useful practices are rarely outlined in such clarity and detail, and the article by Bowers and Voors will be very helpful to experienced and novice researchers alike.

The contributions to this special issue thus highlight that movements towards greater transparency, reproducibility and credibility in empirical research often depend centrally on the mixing of methods—including both quantitative and qualitative approaches. Moreover, practices such as data sharing and public commitments to research protocols can have important benefits across different types of research designs. Nonetheless, our contributors raise important caveats and suggest limitations that we scholars need to heed, as well. In the rest of this introduction, we discuss these issues and situate the papers in this special issue within the contours of the broader push towards transparency in the social sciences.

II. TRANSPARENCY AND REPRODUCIBILITY: ADVANCES AND CHALLENGES

The utility of multiple methods

The move towards experiments in political science—and "design-based research" more generally (Dunning 2012)—has certainly offered important benefits for causal inference. The advantages of random assignment to treatment conditions are widely understood, and we need not review them here.[3]

Yet, the experimental turn has offered several other benefits—especially relative to the previously standard practice in which scholars tended to use multivariate regression and related methods to analyze off-the-shelf observational data sets.[4] Particularly important in our view, the successful use of experiments and related methods often requires close engagement with research contexts. Indeed, experiments and natural experiments usually involve multi-method research and substantial fieldwork—enabling what David Collier refers to as "extracting ideas at close range" (1999, cited in Dunning 2015: 228). Among other advantages, sustained fieldwork presents the opportunity to use process tracing to bolster causal inference (see Bennett and Checkel 2015; Dunning 2015).

This close-range qualitative work can also improve the quality of quantitative data analysis. For example, causal process observation (CPOs) can help with

---

[3] Manipulation of experimental treatments allows researchers to isolate the effect of particular interventions, while random assignment ensures that treatment and control groups are identical up to random error, save for the presence or absence of the intervention. Other assumptions must hold, but manipulation and random assignment offer powerful resources for causal inference.

[4] Here, "observational" means non-experimental—i.e., data that are not produced in connection with an experimental intervention. Cross-national datasets on democracy or development would be one example.

"model validation"—i.e. they can help assess the assumptions of a causal model that is brought to bear on the analysis of a data set. Key hypotheses—such as the supposition that assignment to treatment of one unit does not affect potential outcomes of other units, usually known as SUTVA or the "no interference" assumption—can be supported or disconfirmed. In a natural experiment research design, qualitative pieces of evidence can also help to bolster (or invalidate) the contention that assignment to treatment is *as-if* random. As one of us has put it, "Before a causal hypothesis is formulated and tested quantitatively, a causal model must be defined, and the link from observable variables to the parameters of that model must be posited. Thus, the credibility and validity of the underlying causal model is always at issue" (Dunning 2015: 221). CPOs drawn from qualitative fieldwork and related methods may play a critical role in this validation of causal models.

In particular, process tracing in design-based research can generate a close understanding of the "selection process"—that is, the process by which units end up assigned to treatment and control conditions. This kind of understanding is critical for research aimed at causal inference—including not just natural experiments, but also conventional observational studies—as Maldonado et al. argue compellingly in their contribution to the special issue. These authors show how qualitative methods may support a particular strategy for "matching," or balancing treatment and control units on observed (as well, it is hoped, as unobserved) covariates. Here, Maldonado et al. demonstrate how mixed-methods research is crucial for improving research transparency and the credibility of models in observational as well as experimental research.

Thus, one key merit of the move towards designed-based research is that this kind of research highlights the value of mixing methods for causal inference. To be sure, there are many difficulties involved in the successful use of causal process observations to validate assumptions about causal process. For example, researchers should certainly heed Bennett and Checkel's (2015: 21) advice to "consider the potential biases of evidentiary sources". Dunning (2015) describes the way in which causal-process observations can sometimes lead to faulty inferences in multi-method research, rather than improve causal assessment. He suggests that scholarly "adversarialism" may help to improve the quality and credibility of process tracing. Yet even if doing so is not always sufficient on its own, integrating multiple methods can be critical for improving the quality and credibility of causal inference in the social sciences.

Consolidating pre-analysis plans for varied research approaches

The experimental turn has also promoted the development of tools for public commitment to research protocols, including study registration and pre-analysis plans. In part, this has stemmed from the recognition that while experimental designs might improve the quality of causal inference in a given study, the use

of experiments does not on its own offer any corrective to problems such as "publication bias"—the tendency of journals to publish only non-null results and therefore to reflect a biased sample of actual research results rather than the actual universe of findings in a given research area.[5] As we will discuss momentarily, while scholars have actively debated the merits and demerits of tools such as pre-analysis plans or results-blind review, these practices may offer useful correctives to publication bias and can foster welcome forms of open inquiry across many different styles of research.

However, among scholars who do not employ experimental methodology, transparency requirements—especially pre-registration—might well appear to collide with ethical constraints; hinder the development of creative scholarly work; or end up making it more difficult to get one's work published. These are important concerns and need to be addressed. In this section, we briefly review several recent advances and consider their usefulness as well as possible limitations for qualitative and multi-method research.

As a preliminary, we first discuss several of the tools that have been proposed for combatting publication bias and related ills. It is useful to distinguish between study registration and potentially more binding forms of commitment, including pre-analysis plans. (Other mechanisms for limiting publication bias, such as results-blind review, are discussed below). Study registration refers simply to documenting the existence of a study in advance of its execution. To date, registration has been somewhat ad hoc, with several different organizations providing third-party registration services. Several political science journals now have a policy of encouraging study registration. According to its advocates, registration is important because it allows tracking of the existence of studies, which is critical for establishing the universe of studies and allowing for possible dissemination of null effects. However, registration is typically voluntary, and the level of detail provided about the planned study varies greatly; for example, detailed pre-analysis plans—considered next—may or may not be included. Study registration may therefore assist in documenting the existence of publication bias, but does little to remedy it (Malhotra 2014).

More critical for improving research transparency may be the registration of pre-analysis plans, i.e., recording the protocol in advance of recording or analyzing outcome data. Pre-analysis plans describe the hypotheses and statistical tests that will be conducted. Yet, there is currently no generally accepted standard for their content: empirically, pre-analysis plans involve more or less specificity. At one extreme is Humphreys et al.'s (2011) approach of posting complete analysis code with mock data, which allows analysts to simply run the code once the real outcome data are collected. Dunning et al. (2015) take a related approach, which is to post analysis code with real outcome data but randomly

---

[5] See Gerber, Green, and Nickerson (2001) or Gerber and Malhotra (2008) for evidence of publication bias. See also Dunning (2016) on the inability of experiments to promote effective cumulative learning, absent other practices and procedures.

reshuffled treatment labels. An interesting alternative that could be feasible in some contexts would be to use pilot data to establish the analysis plan, which would then be then registered for the main part of the study. An important issue that scholars analyzing experimental data have discussed is how and when to use amendments to analysis plans, as inductive learning about features of the study site or the nature of the data necessitates design changes.

Pre-analysis plans clearly have an important role to play in boosting the credibility of statistical tests, especially for avoiding the "file drawer" problem in which only some (nominally "significant") results are reported—a practice that wreaks havoc with the interpretation of significance tests. Such "p-hacking" is widely recognized as a problem, and pre-analysis plans are a part of the solution (Simonsohn et al. 2014). However, if publication bias stems from the unwillingness of journal editors or reviewers to publish null results, it is possible that pre-analysis plans will not do much to correct that problem.

Beyond their role in making statistical analysis more credible, however, the idea that public commitment to research hypotheses could play a useful role in other styles of research is also taking hold. Indeed, a vigorous discussion is emerging in the discipline about the types of research for which pre-analysis plans are appropriate; about implications for scholarly creativity and the extent to which exploratory (not pre-registered) analysis should be reported along with registered analyses; and about the usefulness of adaptive plans and the practice of filing amendments (before analysis of outcome data) as features of the particular context become clear during the research process. Yet, the idea that criteria for reaching certain conclusions—for instance, specifying the kinds of process-tracing tests and the evidence that would be considered necessary or sufficient for validating a hypothesis—could be spelled out in advance of corroborating fieldwork appears attractive.[6]

In this issue, Piñeiro and Rosenblatt introduce the rationale and the advantages of PAP for qualitative research, while also clarifying several of the limitations. These authors combine recent advances in the formalization of qualitative research and propose a guide to constructing a PAP-Q. Regardless of whether a PAP-Q is used, these authors argue that it is critical for researchers to adhere to transparency standards. For instance, researchers should demonstrate that they have "cast the net widely for alternative explanations" or been "equally tough on the alternative explanations" (Bennett and Checkel 2015: 21). Greater provision of supporting qualitative documentation improves the analytical value of qualitative research.

Other efforts to limit publication bias could conceivably find parallels in qualitative and multi-method research, as well. Results-blind review, for example, refers to the practice of reviewing a research report blind to the study's

---

[6]   Such systematization of process-tracing tests is in the spirit of recent work by Fairfield (2013) and Piñeiro et al. (2016).

findings. Thus, referees would evaluate a journal submission on the basis of the importance of the research question, the quality of the theory, and the strength of the empirical design—but not the study's findings. Yet, whether such a review process could work for qualitative articles is an open question, and one to which our contributions suggest some interesting potential answers.

The importance of third-party validation

Finally, we turn to the reproducibility of research. This term is broad but certainly encompasses the ability of third parties to reproduce the results of research in a narrow sense—for instance, produce tables or figures included in an article when given access to an original author(s)' data and code. It may also encompass more demanding forms of critical appraisal, including external replication of a study with new data or subjects.

Yet, what policies or practices can make critical appraisal effective? Encouragingly, after many years of effort in this direction, several journals in political science do require replication data for published articles or have some policy establishing an expectation of data availability. These tend to be among the leading journals in the discipline—even if overall, a recent survey of 120 peer-reviewed political science journals found that only 19 even had a replication policy (Gherghina and Katsanidou 2013). However, even the availability of replication data does not guarantee that it can be used effectively: third parties would need access to a full data set, including variables not analyzed in a published article, to assess the possibility of p-hacking. Ideally, this would occur in conjunction with a pre-analysis plan.

Critical appraisal is made even more difficult when—even if replication data sets are nominally available, in the sense that some data exist from which it would be theoretically possible to reproduce tables or figures that appear in a published paper—detailed, commented replication code is not provided. Bowers and Voors provide a really excellent discussion of this issue, with many striking examples of what can go wrong. Even the original authors of a paper may not be able to recreate or defend a complicated set of data analytic decisions, years after the fact. Bowers and Voors provide enlightening suggestions than can make our research more intelligible to and therefore reproducible by not only other scholars, but also to our "future selves."

In the case of qualitative research, several steps have been taken in the last decade to improve third party validation.[7] We discussed above the idea that scholars immersed in a given substantive area might debate the evidentiary quality of process tracing, helping to overcome the problem that the knowledge

---

[7] Main advances and challenges on qualitative data gathering and analysis transparency are presented in Elman, Kapiszewski, and Vinuela (2010), Lupia and Elman (2014), and Büthe et al. (2015).

needed to discover and validate CPOs is often esoteric. In this vein, procedures that facilitate the adversarial logic of CPO validation could be extremely useful. One clear example is to catalogue archival documents and transcripts of qualitative in-depth interviews, and post them in a repository, e.g., in the Qualitative Data Repository at Syracuse University. As Piñeiro and Rosenblatt highlight, qualitative methods have advanced greater research transparency, such as the posting of entire interview transcripts. As Dunning (2015: 232) puts it: "Better use of such transparent research procedures would likely make for better process tracing by individual researchers in several ways. It would encourage them to think through their own argument and it might also help to assess whether they are being 'equally tough on the alternative explanations.'"

Conclusion: Where to go from here?

Improving causal inference, and especially boosting the capacity for cumulative learning from empirical research, is closely connected with transparency standards. When presenting research making causal claims, as in many other forms of research, it is critical for authors to demonstrate how they arrive at conclusions by disclosing sources and procedures. Indeed, this is an integral part of the scientific method. Over recent decades, the social sciences have made progress toward this goal.[8] The recent focus on study registration and related tools can be seen as the latest chapter in this effort. Transparency in design and analysis should be pursued to the greatest possible degree: in the end, the goal is to stimulate informative debates, as well as conclusions that we can trust about important substantive research questions.

Of course, as many authors have claimed before, the stage of the research agenda, the degree of expertise, and the researcher's familiarity with the context in which he is conducting the study may provide valid reasons to question the viability of registering a complete pre-analysis plan, or pursuing the other recommendations of the contributors to this issue of *RCP*. Another main challenge is to change the incentives researchers face and to create the necessary supportive infrastructure.[9] Although there remains much work to be done to engender an environment that facilitates transparency, in recent years we have witnessed significant progress. For example, there now exist various repositories for qualitative data; researchers have various ways to register pre-analysis plans and all necessary documentation; and, critically, there is funding available to conduct replications. Another potential manner of encouraging the development of transparent research environments is to treat study replication as a component of graduate training in methodology.

---

[8]   The publication of King (1995) is one hallmark reference.
[9]   For a more thorough discussion on this point, see Dunning (2016).

The contributions to this special issue therefore suggest that advancing this agenda is feasible—even if it needs to be done with care and in a manner that preserves and amplifies the contributions of mixed-method and qualitative research. Bowers and Voors detail practices to enhance reproducibility, a key aspect of research transparency and of efficient scientific work more generally. Piñeiro and Rosenblatt present a rationale to register pre-analysis plans for qualitative research. Finally, Maldonado et al. suggest practices that improve causal inference in observational studies. We would also point readers towards several helpful tools and guides developed by researchers associated with the Evidence in Governance and Politics (EGAP) network; these include the Research Design Form that is used in connection with study registration and the Declare Design platform, a new R package developed by Blair et al. that provides a framework for characterizing analytically relevant features of research designs (Blair et al. 2016). These and other new tools enhance the transparency and reproducibility of multi-method research and therefore help produce better conclusions about important substantive topics.

REFERENCES

Bennett, Andrew and Jeffrey T. Checkel. 2015. "Process Tracing: from Philosophical Roots to Best Practices". In *Process Tracing. From Metaphor to Analytic Tool*, edited by Jeffrey T. Checkel and Andrew Bennett. New York: Cambridge University Press, 3-38.

Blaire, Graeme, Jasper Cooper, Alexander Coppedge and Macartan Humphreys. 2016. "Declaring and Diagnosing Research Designs". Unpublished manuscript.

Bueno, Natália S., Thad Dunning, and Guadalupe Tunón. 2014. "Design-Based Analysis of Regression Discontinuities: Evidence from an Experimental Benchmark". Available at SSRN 2604710.

Büthe, Tim, Alan M. Jacobs, Erik Bleich, Robert Pekkanen, Marc Trachtenberg, Katherine Cramer, Victor Shih, Sarah Elizabeth Parkinson, Elisabeth Jean Wood, Timothy Pachirat, David Romney, Brandon Stewart, Dustin H. Tingley, Andrew Davison, Carsten Q. Schneider, Claudius Wagemann and Tasha Fairfield. 2015. "Transparency in Qualitative and Multi-Method Research: A Symposium". *Qualitative and Multi-Method Research: Newsletter of the American Political Science Association's QMMR Section* 13(1): 2-64.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press.

Dunning, Thad. 2015. "Improving process tracing. The Case of Multi-Method Research". In *Process Tracing. From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. New York: Cambridge University Press, 211-236.

Dunning, Thad. 2016. "Transparency, Replication, and Cumulative Learning: What Experiments Alone Cannot Achieve". *Annual Review of Political Science* 19: 541-563.

Druckman, James N., Donald P. Green, James H. Kuklinski and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science". *American Political Science Review* 100(4): 627-635.

Elman, Colin, Diana Kapiszewski and Lorena Vinuela. 2010. "Qualitative Data Archiving: Rewards and Challenges". *PS: Political Science & Politics* 43(1): 23-27.

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

Gherghina, Sergiu and Alexia Katsanidou. 2013. "Data Availability in Political Science Journals". *European Political Science* 12(3): 333-349.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: WW Norton.

Gerber, Alan S., Donald P. Green and David Nickerson. 2001. "Testing for Publication Bias in Political Science". *Political Analysis* 9(4): 385-392.

Alan Gerber and Neil Malhotra. 2008. "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals". *Quarterly Journal of Political Science* 3(3): 313-26.

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter van der Windt. 2011. "Social and Economic Impacts of Tuungane 1: Mock Report". Retrieved on 24 July 2015 from http://cu-csds.org/wp-content/uploads/2011/03/20110304-MOCK-REPORT-SHARED-FOR-REGISTRATION.pdf.

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

King, Gary. 1995. "Replication, Replication". *PS: Political Science & Politics* 28(3): 444-452.

Lupia, Arthur and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency". *PS: Political Science & Politics* 47(1): 19-42.

Malhotra, Neil. 2014. "Publication Bias in Political Science: Using TESS Experiments to Unlock the File Drawer". Presented at the *West Coast Experiments Conference*, Claremont Graduate University, 9 May 2014.

Moravcsik, Andrew. 2014. "Transparency: The Revolution in Qualitative Research". *PS: Political Science & Politics* 47(1): 48-53.

Piñeiro, Rafael, Verónica Pérez and Fernando Rosenblatt. 2016. "Pre-Analysis Plan: The Broad Front: A Mass-Based Leftist Party in Latin America: History, Organization and Resilience". Retrieved from: http://egap.org/registration/1989

Simonsohn, Uri, Leif D. Nelson and Joseph P Simmons. 2014. "P-Curve: A Key to the File-Drawer". *Journal of Experimental Psychology: General* 143(2): 534-547.

Thad Dunning is Robson Professor of Political Science at the University of California, Berkeley and directs the Center on the Politics of Development. Email: thad.dunning@berkeley.edu

Fernando Rosenblatt is Assistant Professor, Escuela de Ciencia Política, Universidad Diego Portales, Chile. Email: fernando.rosenblatt@udp.cl