

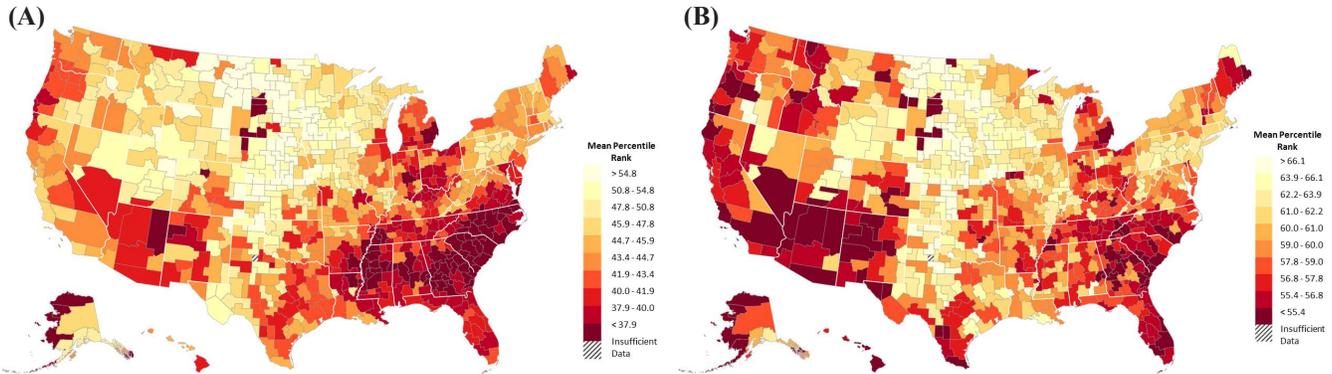
THE POLITICAL ECONOMIST

NEWSLETTER OF THE SECTION ON POLITICAL ECONOMY, AMERICAN POLITICAL SCIENCE ASSOCIATION

Co-Editors:

JOHN AHLQUIST, MEGUMI NAOI, & CHRISTINA SCHNEIDER,
UNIVERSITY OF CALIFORNIA, SAN DIEGO

"Worth a Thousand Words" Neighborhoods and Intergenerational Mobility



This image is from Raj Chetty and Nathaniel Hendren’s “The Effects of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects” (forthcoming in *The Quarterly Journal of Economics*, Oxford University Press). Chetty & Hendren investigate the regional variation in children's economic outcomes as a function of their parents' position in the income distribution. These maps plot children’s mean percentile ranks in the income distribution at age 30 conditional on having parents at the 25th percentile (Panel A) and 75th percentile (Panel B). The maps are constructed by grouping Commuting Zones (CZs) into ten deciles and shading the areas so that lighter colors correspond to higher outcomes for children. Areas with fewer than 10 children, for which we have insufficient data to estimate outcomes, are shaded with the striped pattern. The sample includes all children in the 1980 birth cohort in our analysis sample whose parents are permanent residents (i.e., whose parents do not move across CZs between 1996-2012). To construct these estimates, we first regress children’s family income ranks on a constant and their parents’ family income ranks separately for each CZ and birth cohort. We then define the predicted income rank for children with parents at percentile p in CZ c in birth cohort s as the intercept + p times the slope of this regression. Panel A reports the predicted child rank for parents at $p = 25$, which corresponds to an annual household income of \$30,000. Similarly, Panel B reports the predicted child rank for parents at $p = 75$, which corresponds to an annual household income of \$97,000.

WHAT'S INSIDE THIS ISSUE

<p>"WORTH A THOUSAND WORDS".....1</p> <p>SECTION ORGANIZATION.....2</p>	<p>FROM THE EDITORS.....2 JOHN AHLQUIST, MEGUMI NAOI, & CHRISTINA SCHNEIDER</p> <p>FEATURE ESSAY4 THAD DUNNING, GUY GROSSMAN, MACARTAN HUMPHREYS, SUSAN D. HYDE, & CRAIG MCINTOSH</p>	<p>FEATURE ESSAY.....10 SARAH BUSH, AARON ERLICH, LAUREN PRATHER, & YAEL ZEIRA</p> <p>FEATURE ESSAY.....14 DAVID D. LAITIN</p>
---	---	--

The Political Economist is a publication of the APSA Organized Section on Political Economy. Copyright 2018, American Political Science Association. All rights reserved. Subscriptions are free to members of the APSA Section on Political Economy. All address updates should be sent directly to APSA.

THE POLITICAL ECONOMIST

AMERICAN POLITICAL SCIENCE ASSOCIATION

POLITICAL ECONOMY SECTION OFFICERS

CHAIR

Isabela Mares
Columbia University

SECRETARY/TREASURER

David Primo,
University of Rochester

PROGRAM CHAIRS

Victor Menaldo,
University of Washington
Rory Truex,
Princeton University

EXECUTIVE COUNCIL

Leonardo Arriola,
University of California, Berkeley
Mark Dincecco,
University of Michigan, Ann Arbor
Stephanie Rickard,
London School of Economics
Jonathan Rodden,
Stanford University
Kellee Tsai,
*The Hong Kong University of
Science & Technology*
Rachel Wellhausen,
University of Texas, Austin

NEWSLETTER EDITORS

John Ahlquist,
University of California, San Diego
Megumi Naoi,
University of California, San Diego
Christina Schneider,
University of California, San Diego

NEWSLETTER ASSISTANT

Amanda Harris,
University of California, San Diego

FROM THE EDITORS

The theme for the Spring 2018 issue of *The Political Economist* is “Credibility and Transparency in Political Economy Research: Real Lessons Learned from Actual Research.”

The so-called “credibility revolution” (Angrist & Pischke 2010) in economics and parts of political science has put renewed emphasis on research design considerations and a premium on causal identification. Political economists, especially those working on development, have produced some of the most influential research (and research designs) in this area. At the same time, however, we have seen a ballooning “replication crisis” in medicine, psychology, and economics (Cranmer *et al.* 2016, Ioannidis 2005, Open Science Collaboration 2015). There is no reason to believe that political science and other social sciences are immune.

The conjuncture of revolution and crisis has led to a series of proposals and tools aimed at improving research transparency, replicability, public accessibility, and, ultimately, persuasiveness. These tools range from replication and re-estimation projects in graduate seminars to study pre-registration, meta-analysis, DA-RT, and experiments with results-blind reviews with publication pre-commitments (for example, *Comparative Political Studies* and the *Journal of Development Economics*).

We wanted to find out how political economy researchers are experiencing these tools in their actual, applied work. As a result, this issue departs from the *TPE* tradition of presenting important substantive debates in the form of pithy essays from key interlocutors. Instead, we approached a series of research teams and asked for their critical reflections on their *experiences* with these new research and publication procedures. We posed the following questions: What exactly did you do to improve your research credibility and why? What worked, and what did not? Did you learn anything you would not have learned otherwise? Did you feel stymied or constrained? Were there unforeseen contingencies that led you to revise or abandon your ex-ante plans or the project altogether? What was the publication process like? Would you pursue this strategy again in future projects?

The first group we approached is the EGAP “Metaketa” team, led by **Thad Dunning** (University of California, Berkeley), **Guy Grossman** (University of Pennsylvania), **Macartan Humphreys** (Columbia University and WZB Berlin), **Susan Hyde** (University of California, Berkeley), and **Craig McIntosh** (University of California, San Diego). In our estimation, the Metaketa project is the most developed, ambitious, and extreme example of researchers integrating all the newest credibility-enhancing procedures and tools in a single set of projects for joint analysis and dissemination. We wanted to know how it’s going. Given the project’s scope and ambition, we then allowed the contributing teams to the Metaketa I project on government accountability to briefly summarize their experiences to date.

We also wanted to hear about researchers’ experience with results-blind review and publication pre-commitment. **Sarah Bush** (Temple University), **Aaron Erlich** (McGill University), **Lauren Prather** (University of California, San Diego), and **Yael Zeira** (University of Mississippi) provide critical and insightful reflections on their experience with the *CPS* experiment last year.

Finally, **David Laitin** (Stanford University) poses the provocative question “Is there transparency if no one is looking?”, asking whether all the pre-analysis plans and replication files in the world will amount to much if there are no institutionalized incentives for the research community to actually pay attention to them.

Reading Laitin’s essay in light of the Metaketa and the *CPS* projects provides a fascinating juxtaposition: expanding supply of transparency and credibility, via pre-analysis plans, documentation and replication files, may amount to little if the demand-side incentives to critically engage those materials are absent. The contributors of the

continued on page 3

THE POLITICAL ECONOMIST

From The Editors...continued from page 2

first two essays focused on limitations of their designs from an author's perspective, noting the increase in work intensity, the greater up-front time investment, and the challenge of reacting to unforeseen events. By contrast, David Laitin approaches the problem from the perspective of a critical consumer, whether that be a graduate student learning to digest unfamiliar research or an over-burdened peer reviewer confronting the expanding "supplemental appendices" accompanying every submission.

Laitin highlights that more material does not mean better or more cumulative research if we lack the incentives to dig through it. And digging is what the reader is forced to do under current practice. The static scientific paper, long the dominant form of scholarly discourse, may not be up to the task anymore (Somers 2018). New tools like Markdown, Shiny and Jupyter notebooks can embed the actual data and analysis at a reader's fingertips, producing "living documents."

Finally, we think this issue is timely because the APSA political economy section is actively considering launching its own journal. Given the costs, challenges, and rewards identified here, can the section design a publication process to mitigate the "file-drawer problem" while encouraging critical engagement? Is a traditional journal the best format for achieving this?

References:

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24 (2): 3-30

Camerer, Colin F. *et al.* 2016. "Evaluating replicability of laboratory experiments in economics." *Science*. 351(6280): 1433-36.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False". *PLoS Medicine*. 2 (8): e124

Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science." *Science*. 349(6251): aac4716.

Somers, James. "The Scientific Paper is Obsolete." *The Atlantic*. April 5, 2018.

https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/?single_page=true



FEATURE ESSAY

Reflections on Challenges in Cumulative Learning from the Metaketa Initiative

Thad Dunning (University of California Berkeley), Guy Grossman (University of Pennsylvania), Macartan Humphreys (Columbia University), Susan D. Hyde (University of California, Berkeley), & Craig McIntosh (University of California, San Diego)

The Metaketa Initiative is a new research model aimed at improving knowledge accumulation on topics where real world practitioners and academic researchers share substantive interests. The model, spearheaded by the Evidence in Governance and Politics (EGAP) research network, but with an open call for participation, includes a commitment to fund and coordinate field experimental studies across countries, clustered by theme, in an effort to improve and incentivize innovative research alongside integrated analysis and publication. The “Metaketa” Initiative, meaning accumulation in Basque, is built around close coordination between similar studies and a pre-registered meta-analysis of those studies’ results. Together with seven teams of top-notch researchers, we collaborated on the inaugural Metaketa, which focused on information and accountability.¹

The core question motivating Metaketa I was whether providing citizens with information about politician performance can help facilitate political accountability. All participating studies included a common treatment arm that sought to provide objective information to voters, individually, in the immediate run up to elections. Overall, we expected that voters who received news about incumbents that represented an improvement over their priors would be more likely to turn out and vote for incumbents, and those who received “bad” news relative to their priors were expected to become less likely to vote for the incumbent and turn out to vote. The results were not at all what we expected, though as they are still being finalized, we elected not to report them here. We are confident, however, that the Metaketa model has allowed us to draw conclusions that would have been difficult to draw with confidence from any single study. All of the participating study teams also included a second or innovative treatment arm, as described below, and many teams have working papers or recently published articles in addition to their contributions to the book.²

Overall the experience has been valuable and challenging in equal parts. EGAP is continuing to coordinate the Metaketa Initiative with new clusters of studies on three topics related to natural resource governance, community policing, and taxation. As partial confirmation that the experience is a valuable opportunity, some participants in Metaketa I are now part of new Metaketes.³ Below we discuss the motivation for this initiative, briefly outline the model, and conclude with reflections on the Initiative’s ability to realize its stated goals.

The Challenge of Cumulative Learning

Our starting point was a growing recognition that both theory and policy should not rely on the findings of any one study, since those may not travel across time, space, and study populations. Some findings, however, may generalize beyond the context of a single research study. External validity debates are better evaluated with evidence rather than being left to conjecture.

Meta-analysis seems to offer a reasonable solution to the challenge of generalizability by combining the results of multiple studies on the same topic. Unfortunately, due to existing practices and scholarly career incentives, meta-analysis faces several critical hurdles that constrain its potential as a widespread tool for knowledge accumulation in the social sciences. These difficulties include:

1. **Study sparsity.** Often there are too few studies on the same topic, in part because academics have lower returns to replication relative to “groundbreaking” work.
2. **Study heterogeneity.** Even when scholars conduct studies on related topics, differences in interventions, outcomes, and measurement of inputs and outputs can severely limit the comparability of results—a necessary condition for meta-analysis.

1 Dunning et al. (forthcoming). For the list of projects, see the Metaketa I website. Contributors to the project are: Claire L. Adida, Eric Arias, Taylor C. Boas, Mark Buntaine, Sarah Bush, Simon Chauchard, Anirvan Chowdhury, Clara Correia, Jessica Gottlieb, F. Daniel Hidalgo, Marcus Holmlund, Ryan Jablonski, Eric Kramon, Horacio Larreguy, Malte Lierl, John Marshall, Gwyneth McClendon, Marcus A. Melo, Gareth Nellis, Daniel L. Nielson, Paula Pickering, Melina Platas, Pablo Querubin, Pia Raffler, Catlan Reardon, and Neelanjan Sircar.

2 See, for example, Adida et al. (2017b), Adida et al. (2017a), Boas, Hidalgo and Melo (2017), Arias et al. (2017b), Arias et al. (2017a), Platas and Raffler (2017).

3 Note that the majority of studies in ongoing Metaketes are funded by DFID.

continued on page 5

THE POLITICAL ECONOMIST

Dunning, et al. Feature Essay...continued from page 4

3. **Publication bias.** Too often, referees and editors reject well-designed studies reporting null effects, and authors do not write up null results in the first place. Publication bias limits cumulative learning from a collection of studies on a particular topic: if only significant effects get published, meta-analysis of published studies would give an exaggerated sense of the causal efficacy of particular policies or programs, in addition to other well-documented problems.⁴

4. **Private data.** Researchers are not (always) required to make their data publicly available. This limits third-party researchers' ability to reconstruct results to verify that authors used best practices when analyzing their data, or to use the data in systematic meta-analyses.

These challenges threaten cumulative learning in the social sciences. The tendency to draw from a small number of high-visibility published studies showing large treatment effects leads to a distorted view of the likely effects of interventions. This is true for studies with credible identification strategies. While the “credibility revolution” of the past decade focused primarily on improving research designs at the study-level, the above challenges underscore the importance of building (institutional) strategies that allow us to better validate and aggregate findings across studies.

The “Metaketa” Approach

The Metaketa Initiative was designed to offer one such institutional vehicle. Metaketes seek to generate cumulative evidence by funding similar field experiments across disparate contexts, and coordinating independent project teams around intervention, theory, hypotheses, measurement and estimation strategies. Metaketa thus increases not only the number but also the comparability of studies in a given topic area.⁵ Unlike uncoordinated studies, the data generated by such an effort is more feasibly integrated in an overall meta-analysis. The inaugural Metaketa explored whether and how voters respond to politically salient information on incumbent performance disseminated just prior to elections.

The Metaketa Initiative is based on a number of core pillars, designed to overcome, to the extent possible, challenges both to the reliability of individual studies and, especially, to the credibility of the overall inferences that can be drawn from a set of related studies.

- All Metaketa studies use field experimental research designs, providing a strong foundation for valid inferences about overall effects when aggregating across studies.
- Because any single RCT may have limited external validity, a Metaketa funds multiple studies on a single topic across diverse contexts.
- A Metaketa includes a meta-analysis of the individual studies, resulting in an overall finding that is more likely to be externally valid.
- Diversity of interventions hinder aggregation; instead a Metaketa coordinates research teams on conceptually similar interventions.
- Participating teams commit to measuring the same variables, including key outcome variables, in a similar way. The core principle here is that differences in findings should be attributable primarily to contextual factors and not to differences in measurement.
- A Metaketa coordinates research on a “common” intervention arm (integrated via meta-analysis), but also builds in planned diversity through inclusion in each study of at least one alternative treatment arm, which reflects individual teams' innovations. This approach is designed to incentivize replication while not undermining innovation.
- To minimize challenges related to selective reporting, both individual studies and the meta-analysis commit to using best practices of analytic transparency, including (i) open data and materials; (ii) third-party replication of analyses prior to publication; and (iii) pre-registration of designs and analysis plans.
- To avoid publication bias, researchers commit to be part of an integrated publication of the results of all of the studies.

Making Metaketes Work

Operationalizing these principles was perhaps the major challenge of our work, and it is worth describing some of the challenges

⁴ Humphreys, Sanchez de la Sierra and van der Windt (2013).

⁵ Note that by saying the studies are comparable we do not mean to suggest that they are identical. Rather, allowing interventions to vary while maintaining comparability can be a virtue.

THE POLITICAL ECONOMIST

Dunning, et al. Feature Essay...continued from page 5

we encountered and lessons learned, as well as the limitations of our approach and the trade-offs this form of research entails.

Similarity across studies: One of the most difficult issues involved our attempt to harmonize interventions, to the extent feasible. In practice, what is meant by “similar” is a difficult question, especially given the need to adjust programs to local contexts, and the near impossibility of crafting an “identical” intervention across widely variant electoral practices. A critical issue is therefore the level of generality – or the point on Sartori’s “ladder of abstraction”⁶ – at which experimental interventions can be deemed to provide comparable tests of the same underlying theory. In the case of Metaketa I, all studies are similar to the extent that they disseminate information on incumbents prior to elections, and we also found commonality in the operationalization of “good” and “bad” news about incumbents across studies.⁷ Yet, studies differ with respect to the level of government, type of information, and mode of dissemination, opening the initiative to an obvious line of critique.

Cumulation versus measurement quality: The best locally available outcome measures are not always available in a symmetric way across contexts. For example, aggregate electoral returns – which in principle might be the most reliable measures of vote choice – vary in quality and the level of analysis at which they are available. The common outcome measure in Metaketa I is therefore self-reported vote choice, as it is available in a symmetric way across studies, yet it is not ideal in that it can also be subject to important limitations like social desirability bias.⁸

Pre-specification: Despite best efforts, pre-analysis plans sometimes omit key features of the planned analysis, or make mistakes that are later corrected. While researchers can sometimes file amendments after catching these (ideally before fielding interventions), this is not always possible, and this can raise tricky reporting issues in this type of project. Ample use of supplementary material and online appendices often helps, but third-party replication and open data are critical for evaluating deviations from PAPs or areas in which PAPs were open to multiple interpretations. Individual publications: To encourage participation in the Metaketa Initiative, in addition to their participation in the integrated publication(s), individual project teams were urged to pursue publication of articles, reporting especially (but not exclusively) on the alternative arms of their studies. We hoped that the distinctions between the projects, especially on the second intervention arms, would ease possible competition between participating teams.

Ex post harmonization: Although we built in a mechanism to harmonize strategies *ex ante*, and to commit to a meta-analytic strategy, there was no well developed mechanism to harmonize findings *ex post*. As in any study, teams learned a lot about their particular cases during implementation and developed case-informed hypotheses to account for the outcomes they saw and the conditions under which effects are more or less likely to hold. The Metaketa model would likely be strengthened if there were mechanisms built in to exploit this kind of learning – for example by systematically putting novel hypotheses developed *ex post* in one study to the test in others.

Overall, we believe the Metaketa research model offers an important means of overcoming the problems of study scarcity, study heterogeneity, and selective reporting. The seven studies planned as part of Metaketa I were related experimental studies on the topic of information and electoral accountability, but we wish to be clear that they are far from identical. The level of *ex-ante* coordination allows feasible pooling of separate results into a single meta-analysis and in principle permits better pinpointing of sources of heterogeneity across contexts, yet we attempted to allow researchers sufficient independence to design interventions and pursue hypotheses appropriate to their context and interest. Integrated publication in turn allows us to report the full set of study outcomes regardless of findings. We also had room to explore the sensitivity of our conclusions to departures from the planned analysis – in our case most prominently that one study could not be fielded.

At the same time, the challenges, limitations, and tradeoffs we encountered should also be weighed carefully against the epistemic value of the approach. In deciding whether launching a Metaketa would be appropriate and useful, one might consider the types of priors held by researchers and policy makers that would be overturned by different types of possible results. The

6 Sartori (1970).

7 That is, information in each study was evaluated according to subjects’ prior beliefs, and thus it is possible to distinguish positive from negative information.

8 We take precautions to ensure the reported results in Metaketa I are not driven by this feature of the outcome data.

continued on page 7

THE POLITICAL ECONOMIST

Dunning, et al. Feature Essay...continued from page 6

state of previous evidence on the topic is also important: neither confirming the obvious nor disproving well-replicated findings appear to be the right terrain for Metaketas. The investment in this effort may be most important when one or several studies have produced provocative findings in one context, and the Metaketa can serve as a means of replicating, extending, or reversing the insights of the initial study. It is clearly not a research model appropriate for all research agendas, but we hope that the model can be extended to new topics and improved upon where researchers have the potential to increase coordination with an eye toward cumulative learning.

To facilitate a conversation about the value of these projects, we also asked participating teams to answer three questions: What did you value most? What was most difficult? Would you recommend participation to a friend? Below are lightly edited responses from five of the seven teams.

Reflections from the Research Teams

Benin Team: Claire L. Adida, Jessica Gottlieb, Eric Kramon, Gwyneth McClendon

Our participation in Metaketa I afforded us the opportunity to conduct a national-level field experiment testing key questions about the relationship between information and politician accountability in Benin. We found the experience to be highly valuable and have recommended it to colleagues for the following reasons.

First and foremost, participation in the Metaketa effort meant an institutionalized mechanism for high-quality feedback from the research design stage to the analysis stage from a wide variety of scholars. It pushed us to think more deeply about how we measure performance information, and whether standardization across contexts is really possible, particularly when these contexts vary according to which type of performance information is both salient and measurable.

Second, the combination of a common treatment arm with an alternative treatment arm created incentive compatibility for all participants: the common treatment arm was necessary for the meta-analysis, and the alternative treatment arm allowed each team to develop its own tests of the moderators of the relationship between information and accountability, guaranteeing each group a unique contribution.

The main challenge we faced is likely attributable to the fact that this Metaketa round was organized, by the nature of the questions we were asking, around actual elections. This meant that each team faced hard deadlines, but also that different teams found themselves at very different stages at any given moment. For example, we faced the earliest election in 2015, while other teams didn't have elections until 2016 or even 2017. As a result, our team found itself having to push for coordination early on and adjust to last-minute changes to the standardized survey instrument even as we trained enumerators one week before roll-out.

Still, the experience was overwhelmingly positive. More than just a funding opportunity, the Metaketa effort provided structured, institutionalized feedback at every stage, and its common goal of conducting sound, replicable social science is something we are proud to have been a part of.

Burkina Faso Team: Marcus Holmlund and Malte Lierl

We most valued the opportunity to exchange with other teams throughout the research process, to contribute to meta-analysis and common methodological standards, and the ability to compare results to those from similar studies in real time. Each of these opportunities added value to our research that would have been inaccessible without participation in the Metaketa cluster.

One issue we particularly struggled with were peak workloads. Our smaller team size and our later fieldwork schedule (which was affected by the 2014 popular uprising and the 2015 failed coup d'état in Burkina Faso) made it challenging to adhere to the common deadlines of the Metaketa Initiative.

We also remain uncertain if all the research coming out of a Metaketa cluster stands an equal chance at being published. Each Metaketa study is a well-crafted, extensively reviewed and documented experiment. However, with multiple studies of very similar design, merit, and results, will peer reviewers and journal editors overcome the traditional bias against a perceived lack of novelty?

That said, from the perspective of doing good research and engaging in inspiring intellectual exchange, we would absolutely recommend participation in a Metaketa cluster. From a professional perspective, it remains to be seen if the effort will be rewarded.

Mexico Team: Eric Arias, Horacio Larreguy, John Marshall, and Pablo Querubín

What we valued the most about the Metaketa was the possibility to work (and network!) with an incredible group of both senior and junior scholars. This was an incredible opportunity to receive an unusual amount of high quality feedback on the theory, ex-

continued on page 8

Dunning, et al. Feature Essay...continued from page 7

perimental design and survey instruments. The steering committee played a particularly proactive role both in terms of providing feedback on the individual projects and on coordinating and on coordinating and mediating across teams for the meta-analysis. The fact that this was the first Metaketa also implied that we had to discuss fundamental issues about the idea of cumulation, meta pre-analysis plan and meta-analysis. Being part of this novel initiative was challenging but great.

The most difficult thing was finding the right balance between the incentives of each individual team to differentiate from the other teams (for publication purposes) while at the same time homogenizing measurement and the intervention to contribute to cumulation and the meta-analysis. The fact that most teams had junior, tenure-track faculty in them made the issue of publication incentives even more challenging. The idea of the second treatment arm as the key differentiating feature was very good. In our case, however, the second arm was too closely related to the common treatment arm. We would advise future teams to consider a second treatment arm that is orthogonal to the common arm, such that they can write a separate paper with it even if the common arm does not work.

We learned a great deal from the process and highly recommend it to other scholars. This is especially useful for junior scholars who can benefit from the funding opportunities, close and frequent interaction with senior and like-minded junior scholars in the field, and would benefit from lots of feedback on experimental designs, both substantively and methodologically.

Uganda I Team: Melina Platas and Pia Raffler

For us, the most exciting part of the Information and Accountability Metaketa Initiative was to grapple with methodological and theoretical questions in a dynamic group of scholars with similar research interests and to learn from and with the other country teams. While there are always challenges to aligning preferences across a large group, a major benefit of the series of Metaketa meetings was brainstorming research design, theory, implementation, and troubleshooting, which allowed for deliberation and joint learning – both with regard to the individual studies and the broader goal of devising and testing a new approach to cumulative social science. In our opinion, the most difficult part of the process was identifying a common arm that everyone was excited about, and attempting to make the common arm comparable across studies such that the meta-analysis was meaningful. The latter may be easier to achieve for some questions and interventions than others, which is something to consider for the design of future Metaketes. We believe the Metaketa probably works best when teams meet fairly regularly, as we did in Metaketa I, and where all teams are committed to a common objective in addition to their own individual goals.

Uganda II Team: Mark Buntaine, Sarah Bush, Ryan Jablonski, Daniel L. Nielson, and Paula Pickering

We especially appreciated the opportunity to collaborate at all stages of project design and management with a group of social scientists expert in both field experimental methods and local contexts. The Metaketa PIs strong methods backgrounds, commitment to transparency, and dedication to ethical field research added much to the evolution of our own projects design, pre-analysis plan, and ultimate publications. The multiple collaborative workshops also provided critical feedback and improved the rigor of our project. Finally, we valued the opportunity to contribute to cumulative learning on policy-relevant research through pre-specified meta-analysis and pre-committed joint publication in which the scientific value hinged on design and not on particular findings.

We found it especially challenging to fill gaps in our foresight of specific on-the-ground circumstances and to align our adjustments with the demands of coordination across studies, especially with the hard deadline of the elections looming. Difficulties thus arose in implementing a large and complex field experiment in a semi-democratic developing country on a fixed timeframe while simultaneously providing feedback to and learning from other field experiments in different stages of implementation and analysis. The heterogeneity across research contexts naturally created challenges of coordination, though in the end the diversity of studies strengthened the Metaketes generalizability.

We would recommend participation in a Metaketa for our friends whose work focuses on evidence-based and policy-relevant research and embraces cumulative learning and transparency. We would also urge them not to underestimate the time, effort, and range of skills required for participation. Coordination generates overhead costs, draws on greater expertise, and heightens learning, and these must be weighed carefully – but also in comparison with the costs and benefits of individual research.

References

Adida, Claire L., Jessica Gottlieb, Eric Kramon and Gweneth McClendon. 2017a. “Breaking the Clientelistic Voting Equilibrium: The Joint Importance of Salience and Coordination.” AidData Working Paper Num. 48. Williamsburg, VA: AidData.

Adida, Claire L., Jessica Gottlieb, Eric Kramon and Gwyneth McClendon. 2017b. “Reducing or Reinforcing In-Group Preferences? An

continued on page 9

THE POLITICAL & ECONOMIST

Dunning, et al. Feature Essay...continued from page 8

Experiment on Information and Coethnic Voting.” *Quarterly Journal of Political Science* 12(4):437–477.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubin. 2017a. “Does the Content and Mode of Delivery of Information Matter for Electoral Accountability? Evidence from a Field Experiment in Mexico.” Working Paper.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubin. 2017b. “Priors Rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties?” Working Paper.

Boas, Taylor C., F. Daniel Hidalgo and Marcus A. Melo. 2017. “Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil.” Working Paper.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde and Craig McIntosh, eds. forthcoming. *Metaketa I: Information, Accountability, and Cumulative Learning*. Cambridge University Press.

Humphreys, M., R. Sanchez de la Sierra and P. van der Windt. 2013. “Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration.” *Political Analysis* 21(1):1–20.

Platas, Melina and Pia Raffler. 2017. “Meet the Candidates: Information and Voting Behavior in Primary and General Elections.” Working Paper.

Sartori, Giovanni. 1970. “Concept Misformation in Comparative Politics.” *American Political Science Review* 64(4):1033–1053.



FEATURE ESSAY

Lessons from Results-Blind Review

Sarah Bush (Temple University), Aaron Erlich (McGill University), Lauren Prather (University of California, San Diego), & Yael Zeira (University of Mississippi)

In 2016, we had the opportunity to participate in a unique special issue of *Comparative Political Studies* (*CPS*) devoted to research transparency in comparative politics (Findley et al. 2016). This special issue involved results-blind review, with authors submitting either proposals for research that had not yet been conducted or full papers – minus the results – for research that had already been conducted. According to guest editors Michael Findley, Nathan Jensen, Edmund Malesky, and Thomas Pepinsky, the goal of the special issue was “to assess the potential benefits and costs associated with new models of the publication process by studying how new models can work in practice.”¹

When the call for submissions went out, we were intrigued. We had been planning a study of how political culture, personality cults, and authoritarian imagery help dictators stay in power. Although there is a long tradition of studying these topics within political science, there had been limited experimental research in the area. In the end, we conducted a lab experiment in the United Arab Emirates in which we randomized a subliminal prime of the dictator’s image and assessed its effects on compliance with and support for the regime through a series of laboratory games and survey questions. We believed that the experimental method would provide a valuable complement to previous research because governments and citizens display authoritarian iconography in deliberate ways that make it difficult to identify its effect on outcomes such as compliance and obedience (for a further discussion of these points, see Bush et al. 2016). For reasons we elaborate below, we felt that results-blind review could be advantageous because of the opportunity for both intensive feedback during peer review and for our manuscript to be pre-accepted before going to the field.

The review process for the special issue was somewhat unusual. Initially, we submitted a detailed research design that included a full motivation for the study, theory section, and information about all aspects of the proposed empirical research (including measurement, experimental design, and plans for analyzing the data). We received a decision of “revise and resubmit” plus three in-depth reviews and detailed comments from the editors. Then, after revisions, resubmission, and another round of review, the piece was conditionally accepted. Ordinarily, authors breathe a sigh of relief at that stage. But for us, the work simply proceeded to a new stage, as now we had to execute the experiment. After collecting the data and analyzing it, we submitted our completed article to *CPS* alongside a detailed discussion of how and why our data collection and analysis deviated from our original plans.

The Advantages of Results-Blind Review and Pre-Acceptance

Our overwhelmingly positive experience with the *CPS* special issue derives from two features of the results-blind review process. The first feature is that peer review can take place before data collection. The second feature is that articles can be evaluated for publication and pre-accepted based on the quality of their theory and design without the influence of the ultimate research findings. These features provided numerous benefits as we outline below.

Because peer review took place before we implemented our research design, we were able to get serious feedback on the design from the three reviewers as well as the editors. The reviews were comparable to the reviews we have had in more conventional submissions to journals in terms of the apparent match between reviewers’ expertise and the content of our manuscript. Our experience with the results-free reviewers also resembled our other experiences with reviewers in terms of our efforts to try to respond effectively. However, we appreciated that our reviews were of uniformly high quality in terms of their depth and level of engagement with our manuscript. Our favorable experience with our reviews was not unique. Reflecting on the project as a whole, the *CPS* editors suggested that the “reviews were of higher quality than the average review” (Findley et al. 2016, 1682). Specifically, we received comprehensive feedback on the theory and research design at a higher level than our team has experienced with other projects. As might be obvious given the results-blind nature of the reviews, we did not receive the kinds of reviews that speculate about alternative theories to explain results or that emphasize the results’ importance or lack thereof. Instead, we were asked to discuss our interpretation of potential result profiles (such as uniformly null results) before going to the field.

Receiving high-quality reviews prior to executing our study had at least two benefits. The first benefit was that it allowed us to improve our study’s design in important ways, including by developing additional tests of hypotheses, removing certain

1 See call for proposals here: <http://www.michael-findley.com/research-transparency-in-the-social-sciences.html>.

continued on page 11

Bush, et al. Feature Essay...continued from page 10

tests of mechanisms, and making significant modifications to our experimental treatments. For example, in response to the reviewers' comments, we opted to focus our study on identifying *whether* iconography caused changes in individuals' compliance with and support for the regime rather than explaining *why* it had such an effect. Although there are ways to get feedback on a research design outside of the review process (such as at a specialized workshop), not everyone in the discipline has access to these forums. Moreover, we believe reviewers took more time seriously considering our research design since the manuscript did not contain results.

The second benefit was that we were able to make the changes to our experiment that were suggested by reviewers since we received their feedback *before* implementing the study. For example, following the suggestions of a reviewer and the editors, we added several new outcome measures to measure support for the regime. Our initial measures had primarily concentrated on capturing subjects' compliance with the regime, but our theory also had implications for regime support. The review process helped us identify this weakness in our initial design and correct it. Participants in our study were thus told about three policies that the president of the UAE had endorsed and then were surveyed about their own support for the policies. This type of feedback from peer reviewers about weaknesses in the research design typically comes *after* completing data collection. It can be frustrating to receive since it is often too burdensome – and sometimes impossible – to reimplement a study in response to reviewers' specifications. Or, since reviewers understand those challenges, they may simply avoid giving this feedback altogether.

Finally, because manuscripts were pre-accepted without the research findings, we pursued a research project that we may otherwise not have had the incentive to pursue. Scholars must weigh the costs and benefits of a particular project before going forward. On the one hand, researchers think about the likelihood of publication and benefit to the field or humanity from a given project. On the other hand, scholars must consider the costs of time and other resources.

Assessing a study's likelihood of publication typically includes an evaluation of whether a study's findings are apt to interest future editors and reviewers. Although no project's findings are certain, our particular experiment felt risky. We were confident about our research question, theory, and proposed design. But our plan was to experimentally investigate a phenomenon that – to the best of our knowledge – had not been previously studied in this manner and to use a social science laboratory on the campus of New York University Abu Dhabi (NYUAD) that was brand-new. Furthermore, we anticipated a number of logistical hurdles. Null findings – which we ultimately did have – seemed quite plausible *ex ante*.² Moreover, because our article was pre-accepted before going to the field, we were able to use the pre-acceptance to secure additional funding. For at least one grant, we believe having the pre-acceptance was critical to our success.

Thus, in our view, another key benefit of the results-blind review process and pre-acceptance is that it helps encourage scholars to take on innovative and potentially path-breaking work.

The Disadvantages of Results-Blind Review and Pre-Acceptance

In the Monkey Cage, Ben Ansell and David Samuels raised a host of concerns about administering pre-acceptance from their perspective as *CPS* editors.³ Although they raised important concerns about the complications that pre-registration creates in the *initial* review process, the main disadvantages we experienced pertained to the lack of standardized practices for what occurs *after* the initial review, when a journal has already pre-accepted an article based on pre-registration. Indeed, there still needs to be more clarity about what “conditional acceptance” means in the context of pre-registration and how journals will evaluate changes to pre-accepted designs.

When researchers pre-register their designs *without* pre-acceptance, they often file design or estimation protocol amendments at various stages of the process. Such modifications in design can occur after field-testing an experiment, for example. The American Economic Association's RCT registry and the Open Science Framework's pre-registration mechanism specifically allow such documented modifications.⁴

However, such modifications in the context of pre-acceptance of journal submissions raise important issues that have yet to be worked out. For example, if investigators want or need to make a major design change to an already-accepted study, what happens? Who should decide and on what criteria? Currently, there is no protocol for peer or editorial review of such design

2 In our experiment, the subliminal prime of the dictator's image did not affect compliance with and support for the regime. We note that the other studies in the *CPS* special issue, Hidalgo, Lima, and Cannello (2016) and Huff and Kruszewska (2016), did not have null findings.

3 See <https://www.washingtonpost.com/news/monkey-cage/wp/2016/08/25/heres-what-happens-when-scientists-evaluate-research-without-knowing-the-results/>.

4 See <https://www.socialscienceregistry.org/> and <https://osf.io/k5wns/>.

Bush, et al. Feature Essay...continued from page 11

changes.

Thus, researchers may be in uncharted territory when they have reason to revisit the yet-to-be-implemented design of a pre-accepted study. For example, at one point in Abu Dhabi, we discussed whether to modify our design after pre-testing our subliminal treatment in the lab and concluding that it was unlikely to produce significant average treatment effects. We considered changing to a *supraliminal* treatment, which would have been less naturalistic but might have produced a stronger treatment than a subliminal prime. We decided to keep our original design with the subliminal prime because we felt this research design would provide an informative, and more externally valid, test of the theory and because it was the experiment the *CPS* reviewers and editors had vetted and accepted.

A related issue arises when researchers cannot implement some aspects of a pre-accepted study for unforeseen practical reasons. In many field-based studies, practical challenges are inevitable.⁵ In our case, it proved extremely difficult to recruit the full number of subjects that we had proposed in our submission. As often happens in field-based research, a variety of challenges arose, from the NYUAD annual unannounced fire alarm that occurred during one of our largest lab sessions to a traffic accident involving subjects coming to the laboratory from elsewhere in the country. During the time that we had been allotted in the lab and that members from our team were able to spend in the UAE, we were able to recruit enough subjects to reach 82% of the proposed sample (Bush et al. 2016, 1717). Power analyses indicated that reaching the full proposed sample would not have produced significantly different results, and so we proposed stopping data collection. We shared this information with the editors, who supported our proposed way forward. However, this process was stressful for us since we had been given a specific deadline by *CPS* to submit our completed article in order to meet their publication timeline, and we felt that a publication was on the line. We had to make some of these decisions quickly, and it is not clear what we would have done if the editors had felt it necessary to return to the reviewers to weigh in on our proposal to stop data collection.

Two final concerns relate to how full papers are reviewed after researchers include the results and the overall burden on reviewers. With regards to the former, do the editors evaluate the final product, or is it the job of reviewers? How closely will they read the final product? As noted above, at *CPS*, the judgment was carried out by the editors. This strategy gives editors more discretion -- and also more work. For some papers, perhaps those with more technical elements or with more significant proposed deviations, reviewers should potentially carry out this process instead of editors. And for all papers, external review of the final product may be worthwhile, even if papers have received conditional acceptance.

With regards to the latter concern, adding in this layer of additional review would put more onus on reviewers, as does asking them to carefully vet pre-analysis plans. One of us (Erich) has also run a pre-acceptance competition for an area studies journal (*Caucasus Survey*). In that competition, the reviewers noted that, particularly for pre-acceptance based on non-released observational data (such as the pre-acceptance competition carried out by the American National Election Studies, <https://www.erp2016.com/>), the amount of work in the review process was much greater; because the analysis has not been conducted, there is much more room to suggest large scale changes to theory. Therefore, it is harder to judge the standard for article acceptance. This additional experience with a pre-acceptance competition reinforces the idea that the discipline may need to reward reviewers for the larger amount of work that this type of review entails because the incentives to participate are unclear.

Implications

Given the benefits to authors and interest by many scholars, the question of what journals ought to do with regards to pre-acceptance arises. We think that it is too early in the process to proceed prescriptively, but we would encourage journals to “let 1,000 flowers bloom.” We need more experiences first to make strong recommendations. One policy proposal would be for journals to try different types of processes, so that as a discipline we can comparatively evaluate them.

We also believe that experimenting with pre-acceptance mechanisms for non-experimental and qualitative studies based on interviews and focus groups should also be explored. However, first, more work needs to be done to create accepted mechanisms for pre-registering observational and qualitative research (Miguel et al. 2014). At the APSA roundtable we organized on this topic, there were concerns about how reviewers and editors should evaluate submissions in which data had already been collected and could have, in principle, been previously analyzed. The academic review process is not designed to evaluate fraud, and we do not think we should treat pre-registration differently. We believe that if scholars state they have not analyzed the data, they

⁵ In extreme cases, researchers may not be able to implement a pre-registered or pre-accepted study at all. For example, one of the seven pre-registered studies planned as part of an Experiments in Governance and Politics (EGAP) Metaketa was forced to stop in the middle of implementation because of backlash by local politicians (Sircar and Chauchard forthcoming). Other studies also faced smaller-scale implementation challenges. Not being able to implement a study as planned, or even at all, poses important questions about what journals should do in those cases with pre-accepted studies as well as pre-registered meta-analyses.

continued on page 13

Bush, et al. Feature Essay...continued from page 12

should be taken at their word during the review process.

Concluding Thoughts

As explained above, results-blind review enabled us to improve our study through engagement with excellent reviewers and editors and encouraged us to pursue what we viewed as a risky study. It also introduced certain unique stresses related to wanting to be sure to that we carried out the study in such a way that reflected our original protocol even in the face of unanticipated logistical challenges. What it did not change was how we reported our results, although we did feel liberated from any worries about whether null findings would make it difficult for the study to be published.

On a more personal note, the results-blind review process was advantageous for a multi-author collaboration such as ours. Our team of four often had vigorous debates about the direction and design of the project before pre-acceptance. For us, pre-acceptance provided a focal point around which we could coordinate. Moreover, we all felt particularly enthusiastic about carrying out the research knowing the article would be well-placed upon completion. While we cannot say every researcher or team will share our experience – indeed, we could imagine for some pre-acceptance could lead to satisficing or leisurely execution – for us, it was a motivating force.

Ultimately, we had a positive experience with results-blind review and believe that the advantages outweighed the disadvantages. With time, we believe that some of the uncertainties and challenges about how best to implement this process will be addressed. It is likely that results-free review will not and perhaps should not become the dominant type of peer review in our discipline, but we hope that more journals will consider adopting it, at least some of the time.

References

Bush, Sarah Sunn, Aaron Erlich, Lauren Prather, and Yael Zeira. 2016. “The Effects of Authoritarian Iconography: An Experimental Test.” *Comparative Political Studies* 49(13): 1704-1738.

Franco, Annie, Neil Malhotra, and Gabor Simonvitis. 2014. “Publication Bias in the Social Sciences: Unlocking the File Drawer.” *Science* 345(6203): 1502-1505.

Findley, Michael G., Nathan M. Jensen, Edmund J. Malesky, and Thomas B. Pepinsky. 2016. “Can Results-Free Review Reduce Publication Bias? The Results and Implications of a Pilot Study.” *Comparative Political Studies* 49(13): 1667-1703.

Hidalgo, F. Daniel, Júlio Cannello, and Renato Lima-de-Oliveira. 2016. “Can Politicians Police Themselves? Natural Experimental Evidence from Brazil’s Audit Courts.” *Comparative Political Studies* 49(13): 1739-1773.

Huff, Connor and Dominika Kruszewska. 2016. “Banners, Barricades, and Bombs: The Tactical Choices of Social Movements and Public Opinion.” *Comparative Political Studies* 49(13): 1774-1808.

Miguel, E. et al. 2014. “Promoting Transparency in Social Science Research.” *Science* 343(6166): 30–31.

Sircar, Neelanjan and Simon Chauchard. Forthcoming. “Dilemmas and Challenges of Citizen Information Campaigns: Lessons from a Failed Experiment in India.” In *Metaketa I: The Limits of Electoral Accountability*, eds. Thad Dunning, Guy Grossman, Macartan Humphreys, Susan D. Hyde, and Craig McIntosh. Cambridge: Cambridge University Press.

FEATURE ESSAY

Is there transparency if no one is looking?

David D. Laitin (Stanford University)

As a doctoral student in the late 1960s at UC Berkeley, I believed in science. For my dissertation, I had a topic, viz. the failure of the newly independent African states, through their continued reliance on European languages as media of instruction and administration, to fully implement the cultural component of their anti-colonial agenda. The research question that followed was whether there were cultural implications, and not merely learning costs, for the official retention of colonial languages. From having served as a Peace Corps Volunteer in Somali (where I attained conversational competence in Somali), I had strong intuitions about those implications.

But as a budding scientist, intuitions belonged in what we today call a pre-analysis plan. And this would be followed, again in today's lingo, by a field experiment. (I found an advisor in the psychology department to guide me in running experiments, as none of my advisors had seen this done in our discipline). For my field experiment, I trekked to a secondary school in Wajir, in what was then called Kenya's Northeastern Province, where all students were bi-lingual in English and Somali. I designed a set of role-playing scenarios for pairs of students to motivate dialogue on political issues. For each dialogue that was to ensue, I had predictions of the way Somalis would frame the issue when speaking in English and how it would be differently framed when speaking in Somali. In the field, I assigned Somali pairs the language in which the scenario was presented and how it would be discussed. (Alas, I assigned language sequentially, but not randomly). I taped all the dialogues, had them translated and transcribed, and then ran statistical tests to see if there were systematic differences in issue framing depending on the assigned language of the dialogue. There were, and along lines that I had predicted!

But as I was writing up my results, I worried that science was not being served. There was no archived and dated pre-analysis plan. No one but myself coded the transcripts; no one checked my statistical tests; and no one had seen if my results conformed to my initial hypotheses. I copied my notes, my (translated) transcripts, and coding rules and gave this package to my dissertation advisors, expressing my worries that there were no checks on my scientific findings. I heard nothing back from them, but that did not deter them from signing my dissertation. The following year I submitted my manuscript to University of Chicago Press, ready to provide all the raw data. Neither of the peer reviewers seemed concerned as they recommended it for publication. The book appeared as *Politics, Language and Thought: The Somali Experience* (1977).

My greatest hope for the book was that it would require all political scientists writing about the challenges of political development to have a footnote to my work on the cultural costs of ignoring African languages for the new states. This is how a scientific finding was to add to the literature. For the decade subsequent to publication, my scientific finding never got such a citation. Worse, no political scientist sought to undermine my finding, or even to delimit its scope with replications from other languages or countries. Meanwhile, my career advanced just fine, even if I was not part of a scientific community seeking cumulative advances in the field of political development.

A brilliant review of my research program provided a much more optimistic evaluation of its role in scientific progress.¹ In it, Kanchan Chandra argued that the criterion of progress for comparative politics should not be the amount of variance explained by a proposed independent variable, but rather the identification of the operating mechanisms in a well-specified political domain. Chandra focused her commentary not on the cultural implications of reliance on colonial languages (for which I did the field experiment), but rather on the question of why African leaders did not take advantage of independence to further the cultural project that inspired their national aspirations (for which I analyzed bureaucratic interests in the maintenance of the colonial language regime). For Chandra, my argument about the failure to adopt indigenous languages in post-colonial Africa involved a chain of mechanisms, one of which was what I later called "the private subversion of a public good" in which Africans (and in later papers South Asians and Catalans), and especially those inheriting high positions in the civil service, would herald the importance of indigenous languages for their countries. But if an indigenous language gained official status, they (as well as less well-endowed citizens) would borrow money if necessary to support their children in private education to acquire facility in the former colonial language. If subversion is general, the public good is undermined.² Chandra judged this finding not by any prediction on which post-colonial societies will adopt indigenous languages as official (with a focus on explaining outcomes), but rather by which other scholars can follow the chain from the recognition of the incentives to subvert (with a focus on mechanisms). While my hat goes off to Chandra to demonstrate that my work did not fall into replication oblivion, I

1 Kanchan Chandra (2006) Mechanisms and Outcomes, Qualitative Methods Newsletter, American Political Science Association, Spring Issue.

2 For my first elaboration of the "private subversion" mechanism see Laitin (1993).

continued on page 15

THE POLITICAL ECONOMIST

Laitin Feature Essay...continued from page 14

still worry that rigorous testing of reported findings in comparative politics remains underutilized.

Returning to today's lingo, I believed in transparency, replication, and the cumulative development of science. What I learned is that there is no transparency if no one is looking.

Times of course have changed. As a discipline we can be proud that we now promote formal pre-analysis plans for experimental work and batch files for replication. But we still have a long route ahead. Incentives for scholars at the individual level are not aligned with efficient scientific advance. Many issues remain unresolved.³ Standards for pre-analysis plans have not been established. Can they list every possible branch of statistical tests – e.g. if the test of our treatment in sub-group x reveals non-significant results, we then test on sub-group y – *ad infinitum*? If so, this is merely *ex ante* fishing. How many branches can be proposed before peer reviewers are lulled into sleep or refuse to do the work demanded of them? What are the returns to replication, especially if publication of replication results (especially if the replication results are confirmatory) don't have prestigious publication outlets? What are the standards for a publishable replication that adds to or subtracts from our confidence in a theory?

In sum, all the pre-analysis plans and batch files we produce do not serve science if no one has a career interest in deciphering them or confirming the results that followed from them. We have increased the supply of transparency, but have given insufficient attention to generating a demand for it.

References

Laitin, David D. 1993. "The Game Theory of Language Régimes." *International Political Science Review*, 14(3): 227-40.

Laitin, David D. 2013. "Fisheries Management." *Political Analysis*, 21(1): 42-47.

³ I have raised several of these issues in Laitin (2013).

A

P

S

A