



# The Experimental Political Scientist



## In this issue

- Letter from President
- Ostrom on the Role of Experiments
- Moore on Block Randomization
- Lupia on Procedural Transparency
- Druckman on Experimental Myths
- Michelitch on GOTV in Africa
- Announcements

## From the Editor

Welcome to the first issue of the *Experimental Political Scientist*. We have a set of stimulating contributions and more in the pipeline. Our newsletter is driven by its consumer. I welcome submissions on a variety of topics of interest to experimentalists. These can be reviews, short summaries of new methods, suggestions for improving experiments, think pieces,... Also, if you have announcements please send them in and we will publish them. We aim to publish the newsletter in October and May. Happy Experimentation!

## Dustin Tingley, Harvard University Information on Joining or Contributing

The *Experimental Political Scientist* is the official newsletter of APSA Organized Experiments section 42. To receive the newsletters register for the section (\$8/yr!) at <http://www.apsanet.org> and visit us at <http://ps-experiments.ucr.edu/group>. Scholars wishing to contribute to the newsletter should email the editor at [dtingley@gov.harvard.edu](mailto:dtingley@gov.harvard.edu).

## Letter from the President

The formation of the Organized Section on Experimental Research marks an important milestone in the growth and development of political science. Over the past two decades, the discipline has witnessed a dramatic transformation in the prominence of experimental research. Survey, laboratory, and field experiments used to be peripheral research strategies; now, they are central. Their increasing role in the discipline may be measured in many ways, such as the remarkable growth of journal articles using experimental data or discussing the logic of experimentation. The advent of experimental research is part of a broader credibility revolution in the social sciences, a movement that places special emphasis on extracting secure causal inferences.



The Experimental Research Section is founded in a spirit of collaboration that spans subfields and disciplines. The imagination that fuels experimental research flows from substantive expertise of scholars across political science and the social sciences more generally. The new Section provides a forum for scholarly exchange about experimental design and analysis, helping researchers translate their substantive questions into workable and informative experiments. We anticipate that many if not most of the panels hosted by Experimental Research will be co-sponsored with other sections. The formation of the new Section is also an opportunity to take a fresh look at the discipline's core scientific institutions and values. Unlike observational data, which are often gathered by large scale collective ventures such as the World Values Survey, experimental data typically grow out of a researcher's own primary data collection efforts. The documentation and preservation of experimental data require special attention. The same may be said for the format of journal articles. Social science journal articles fall somewhere between the turgid law review essay and the terse physics paper. The Experimental Research Section may wish to consider the question of whether introduce a new journal format, perhaps one that embraces new norms about reporting essential information about experimental design and results. More generally, the Section provides an important voice for scientific norms that suffuse experimental research: transparency and reproducibility.

The fledgling Experimental Research Section depends on the active participation of its members in order to establish itself. There are many ways to get involved. The simplest and most direct way is to join the Section (log into "My APSA" at <http://www.apsanet.org>, click on "Join Organized Sections" and then follow instructions to join Section 42) and attend its sponsored panels at APSA. As you may know, the Section is now in its probationary period and must show steady progress in terms of membership in order to become a Section in good standing. By the same token, the Section is currently allotted the minimum number of panels; to grow that number, high levels of panel attendance are required. Please also let the officers know if you are interested in serving on committees, writing for the newsletter, or becoming an officer. We welcome you aboard.

## A Note for the First Newsletter of the APSA Experimental Section

Elinor Ostrom, Distinguished Professor and Professor of Political Science, Indiana University

Congratulations!

I am very pleased to see the establishment of the APSA Experimental Section. It was a joy to attend the first meeting and see the large number of outstanding political scientists attending that meeting. I look forward to participating actively in the section.

When I first started doing experiments in the early 1980s, many political scientists criticized me for doing experiments. At that time, there was a deep worry about the lack of external validity from the experimental method combined with skepticism that you could learn much from experiments with undergraduate students. When I was president of APSA in 1997, I was asked to give talks at all regional meetings. I used this “bully pulpit” as an opportunity to discuss exciting experimental results and how they provided a foundation for a behavioral approach to the theory of collective action. I reflected on this in my Presidential Address (Ostrom, 1998).

Fortunately over time, two major developments have happened. First, more and more political scientists have started to do outstanding experimental work and are finding important patterns of great value to our discipline. Secondly, the experimental method has been subjected to a variety of tests.

In the 1990s, Juan-Camilo Cardenas took designs that had been run in labs with undergraduate students in American universities and conducted them in their countries. Cardenas translated the experimental instructions we had used in a series of experiments in Bloomington into Spanish. He pretested them to make sure that while the same underlying mathematical structure was retained, the language was meaningful to campesinos in the Colombian countryside. Cardenas (2000, 2004) found many of the patterns of results that we had obtained in Bloomington, Indiana, were replicated by his very carefully designed experiments. Now there are many efforts to do field experiments (Prediger et al., 2010; Velez et al., 2010; Volland, 2008).

Further research has shown that behavior observed in field laboratories is consistent with patterns observed in the

field. For example, Fehr and Leibbrandt (2008) have conducted experiments where they could identify fishermen who were more cooperative in field experiments. Then they measured the nets that the fishermen were using and found that the more cooperative fishermen tended to use fishing nets with a larger mesh size. Thus, a cooperative strategy in an experiment was correlated with cooperative behavior in the field that enabled more fish to escape from a net and replenish a fishery.

In Bloomington, we started a series of common-pool dilemma experiments that were conducted with graduate students and faculty attending a summer institute. When asked to teach an advanced seminar in Bratislava, I used the same design. Over time, we had 189 participants from 41 countries, none of whom were undergraduates. The findings across settings were quite consistent with one another and consistent with experiments run with undergraduates at an earlier juncture (Ahn et al., 2010).

Rustagi et al., in a forthcoming article in *Science*, have conducted a very large number of field studies in Ethiopia where survey research was used to assess the orientation of over 650 individuals coming from 49 user groups sharing forest resources. The three types of data fit well. Those who responded to survey questions showing that they were conditional cooperators (as contrasted to free-riders) tended to behave in the experiments in a way consistent with their survey responses. Further, in communities represented by more individuals who were conditional cooperators, measurements of their forests showed improved conditions as contrasted to villages containing a larger number of free-riders and low performance on the experiment. As more scholars are able to conduct studies that relate findings in the lab to findings in the field, questions related to the validity of experimental outcomes will be addressed.

I do not want to overemphasize the use of experiments in political science more than other methods. We do need to conduct individual case studies over long historical periods. We need meta-analysis of large numbers of individual cases. We need large-N studies. We need to do game theory and agent-based models. Experimental research can be a major complement to the careful use of all rigorous methods (Poteete et al., 2010). What is exciting is that political scientists are now accepting experimental methods as one of our core methods. This newsletter and the new APSA section will help to ensure our continuing along a productive path.

## References

Ahn, T. K., Elinor Ostrom, and James Walker. 2010. "A Common-Pool Resource Experiment with Postgraduate Subjects from 41 Countries." *Ecological Economics* 69(12) (October): 2624-33.

Cardenas, Juan-Camilo. 2000. "How Do Groups Solve Local Commons Dilemmas? Lessons from Experimental Economics in the Field." *Environment, Development and Sustainability* 2(3-4): 305-22.

Cardenas, Juan-Camilo. 2004. "Norms from Outside and from Inside: An Experimental Analysis on the Governance of Local Ecosystems." *Forest Policy and Economics* 6(3-4): 229-41.

Fehr, Ernst, and Andreas Leibbrandt. 2008. "Cooperativeness and Impatience in the Tragedy of the Commons." *IZA Discussion Paper* 3625.

Ostrom, Elinor. 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action." *American Political Science Review* 92(1) (March): 1-22.

Poteete, Amy, Marco Janssen, and Elinor Ostrom. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, NJ: Princeton University Press

Prediger, Sebastian, Bjrn Vollan, and Markus Frlich. 2010. "The Impact of Culture and Ecology on Cooperation in a Common-Pool Resource Experiment." *Ecological Economics*, doi:10.1016/j.ecolecon.2010.08.01.

Rustagi, Devesh, Stefanie Engel, and Michael Postfield. 2010. "Conditional Cooperation and Costly Monitoring Explains Success in Forest Commons Management." *Science*, forthcoming.

Velez, Maria A., James J. Murphy, and John K. Stranlund. 2010. "Centralized and Decentralized Management of Local Common Resources in the Developing World: Experimental Evidence from Fishing Communities in Colombia." *Economic Inquiry* 48(2): 254-65.

Vollan, Bjrn. 2008. "Socio-Ecological Explanations for Crowding-Out Effects from Economic Experiments in Southern Africa." *Ecological Economics* 67(4): 560-73.

## Blocking Political Science Experiments: Why, How, and Then What?

Ryan Moore, Assistant Professor of Political Science,  
Washington University at St. Louis

As experimentalists, we enjoy that moment right before we push the "randomize" button. We have carefully defined

our sample, treatments, implementation protocol, and eventual analysis plans. We're about to overcome all the maladies of non-random treatment assignment that led us to the hard work of setting up an experiment in the first place.

But wait – what will you do if all the poor neighborhoods end up in the treatment condition? What if all the male subjects are assigned to control? What if you have not just one binary measure, but a half-dozen continuous covariates that you want balanced in your finite sample? Bigger samples and good luck will help, but we can do better. *Blocking* the sample prior to randomization can incorporate rich covariate information to ensure comparable groups, increase the efficiency of treatment estimates, and provide guidance should things go wrong.

Blocking is the pre-randomization sorting of units into homogeneous groups with the plan to randomize within those groups. In the examples above, you could sort neighborhoods by income, or subjects by sex, and then randomize treatment assignment within these blocks.

Creating blocks helps ensure that covariates are balanced across the treatment conditions. Consider a small GOTV experiment with six voters who have voted 2, 2, 3, 3, 4, and 4 times in the last four elections. If we randomly allocate half the voters to treatment and half to control, then in 60% of possible randomizations, our two groups will differ in mean previous votes by  $\frac{2}{3}$  or  $\frac{4}{3}$ . However, if we block exactly on the number of previous votes  $X$ , we will always have perfect balance across the treatment conditions. This balance reduces the bias in causal estimates that comes from comparing a treatment group of 2, 2, 3 with a control group of 3, 4, 4, for example.

Blocking also increases the efficiency of causal estimates; this means fewer observations are needed to detect a given treatment effect, saving time and money. Suppose that the outcome is whether a voter votes in this election, voters' baseline probability of turning out is  $0.2X \pm .05$ , and the GOTV prompt increases the probability of turnout by 0.1. Then, the standard deviation (SD) of the difference in treatment and control means from all the unblocked randomizations is about 0.15. Blocking this experiment on  $X$  yields an SD of mean differences of about 0.04 – a design that is about 73% more efficient!

Through blocking, design can anticipate and overcome a frequent field experiment reality: some units may be compromised during an experiment, and they and their

blockmates can be excluded from analysis without endangering the entire randomization.

To implement blocking in an actual experiment, the first decision is to choose the variables to block on. You can block on a large set of covariates, including discrete and continuous measures. Blocking should focus on variables likely to affect the outcome of interest. Similarly, for any important subgroup analyses you have planned, block on the variables that define the subgroups to ensure that enough units from each subgroup are assigned to the various treatment conditions.

Next decide how to weight the blocking variables. Typically, you'll first note firm restrictions you want to place on blockmates. For example, you may randomize polling places *within* metropolitan areas or undergraduate subjects *within* universities. Further, you might want to restrict blockmates to be within a range of one another on a more continuous measure, such as no more than 100 points different in SAT scores. Using the sample data `x100` provided in the R library `blockTools` (Moore 2010), you can, e.g., use the `block` command to block on two continuous variables `b1` and `b2` within groups defined by variable `g`, and restrict blockmates to be no more than 100 points different on `b2`:

```
> out <- block(x100, groups="g",
id.vars="id", block.vars=c("b1", "b2"),
valid.var="b2", valid.range=c(0,100))
```

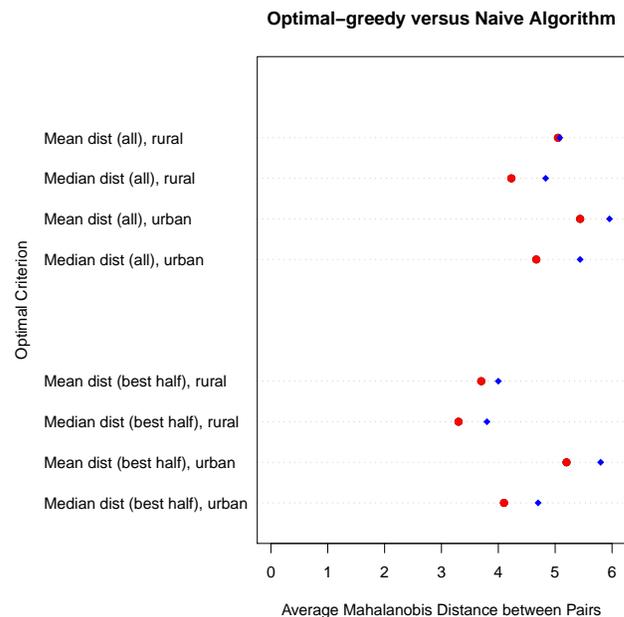
By default, the blocking variables are weighted by the inverse of their covariance matrix using the Mahalanobis distance. If there are outlying observations in  $X$  that you still want to include in the experiment, you can use estimates of the covariance matrix that are robust to these observations.<sup>1</sup> Alternatively, you can exploit substantive knowledge to weight important quantities more highly in the distance calculation.

Finally, you will select an algorithm for creating the blocks. While a naive greedy algorithm will create the best block using the first unit in the dataset, then the second, etc., this blocked design may not be the best design possible. An optimal algorithm considers all possible blockings and selects the one that gives the best balance, but can be computationally intensive even in “medium-sized” samples.

<sup>1</sup>This and all other options mentioned here are available in `blockTools`. For a more full tutorial, including how to install the package, see <http://rtm.wustl.edu/software.blockTools.htm>

<sup>2</sup>The eight comparisons represent two types of units (urban/rural), two subsets of units (all/best half), and two global measures of optimality (mean/median distance).

A middle approach, an “optimal-greedy” algorithm, considers all the multivariate distances between units at once, and selects the best available block. The optimal-greedy approach outperforms the naive greedy algorithm in balancing covariates, and the Figure below shows evidence of this outperformance from the actual field experimental design described in King et al. (2007). The red dots show the decrease in covariate imbalance when compared to the blue dots in several cases.<sup>2</sup>



You now have a table of blocks, ready for random assignment. Using the output object from above:

```
> assg <- assignment(out)
```

Another feature enables you to diagnose potential interference between units by checking whether treatment and control are “too close” to one another. Here we check for units of opposite treatment condition within five points different from one another on `b1`:

```
> diagnose(assg, x100, id.vars = "id",
suspect.var = "b1", suspect.range = c(0,5))
```

After you have implemented your experimental protocol and collected your outcome and follow-up data, you're ready to analyze the blocked experiment to calculate treatment

effects. Typical difference-of-means estimators still apply (see Imai et al. (2009) for related work on cluster randomizations), and parametric regression estimators should include indicators for blocks.

Blocking can help you satisfy scientific colleagues (with less biased estimates), funders (with more efficient design), and policy implementers alike (with plans for compromised units). Never again will you need to worry that your digital coin might misbehave!

### References

Imai, Kosuke, Gary King and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science* 24 (1):29-53.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. 2007. "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management* 26(3):479-509.

Moore, Ryan T. 2010. "blockTools: Block, Assign, and Diagnose Potential Interference in Randomized Experiments." R package version 0.5-1.  
<http://CRAN.R-project.org/package=blockTools>

## Procedural Transparency, Experiments and the Credibility of Political Science

Arthur Lupia, Hal R. Varian Collegiate Professor of Political Science, University of Michigan

Political science has a problem. Your actions as an experimental political scientist can be part of the solution.

The problem pertains to our credibility. Many political scientists publish empirical research claims that others cannot replicate. This outcome occurs even in cases where the scholar who attempts the replication possesses the same dataset as the scholar who made the original claim. Consider, as an example, a situation in which both scholars

have equal access to a public data source, such as the American National Election Studies, but one cannot reproduce the other's published claims.

Across North America and Western Europe graduate-level classes in Political Science try to reproduce empirical claims made in our discipline's leading journals. I have spoken to scholars who teach these classes. The typical reported success rate is abysmal. This is embarrassing for the discipline.

When one scholar cannot reproduce another's empirical claims, particularly when they share access to a common dataset, the failures call into question the credibility of the initial claims. Credibility is called into question because it is often difficult to separate the meaning of an empirical claim from the processes that produced it. In other words, the meaning of the claim "If X, then Y," often depends on how X and Y are measured and on how the relationship is examined.

When scholars cannot recall, or find a record of, the steps they took in producing an empirical claim, then they are handicapped in their ability to render a credible explanation of what their result means. For example, when a scholar manipulates ANES variables in ways that he or she fails to record and/or cannot remember which specific regression model produced the results in his or her paper, readers are justified in questioning the initial claim's meaning. While experimental scholars are likely familiar with such problems in quantitative Political Science, they are also manifest in qualitative scholarship (see, e.g., Moravcsik 2010).

Current and future leaders of experimental political science have a unique opportunity to make a difference in the domain of procedural transparency. In the opening years of our organized section, we have an opportunity to establish best practices for documenting and sharing information about our procedures. In this essay, I will offer suggestions about the practices we should pursue and argue that if experimental political scientists commit to high and consistent levels of procedural transparency, the cumulative effect of such commitments will be to improve Political Science's credibility.

### *How Procedural Transparency Increases Credibility*

The goal of this essay is to encourage experimental political scientists to augment their individual, and the field's collective, credibility by committing to high levels of

procedural transparency. A few definitions can clarify the goal. By credibility, I mean the quality of being believable or trustworthy. Many political scientists are already working hard to offer credible explanations of a range of important phenomena. One way that a political scientist can gain credibility is to derive her or his conclusions by means that scholarly audiences see as legitimate. By legitimate, I mean that actions are taken in accordance with recognized or accepted standards or principles.

Part of what makes political science credible is that its practitioners take actions in accordance with widely accepted views of the scientific method. For the purpose of this essay, I will emphasize two components of the scientific method. First, the method entails the collection of data through observation and experimentation and the formulation and testing of hypotheses. Second, there is a basic expectation to document, archive, and share all data and methodology so that they are available for careful scrutiny by other subjects.

In political science, these attributes of the scientific method are implemented in various ways. While some scholars build theories that produce testable hypotheses but do not test the hypotheses themselves, others focus more on hypothesis testing than theory development. All have important contributions to make. Moreover, all such scholars have the opportunity to make their methods available for careful scrutiny as a means of showing that they are acting in accordance with accepted standards or principles. Such transparency can provide evidence of legitimacy and can give people a reason to believe that published results are relevant and meaningful to political phenomena. Procedural transparency can increase credibility.

In the domain of political science, however, a desire for procedural transparency can be constrained by privacy concerns. In some cases, the information collected about research subjects poses a potential threat to their privacy. For example, the American National Election Studies collects detailed information about where respondents live, as such information is necessary to conduct a large set of face-to-face interviews in the six to eight weeks before Election Day. If participants believed that we should not protect their anonymity, they may provide different information (or act in different ways) than they do with current protections in place.

In many disciplines now, pushes for greater transparency

are generating debates about the kinds of evidence that researchers should and should not be expected to share. While experimental political scientists should participate in such debates, such venues are not the place where I feel that our comparative advantage in boosting political science's credibility is greatest. Hence, in the remainder of this essay, I want to focus not on questions about data sharing, but rather on documenting and sharing information about the procedures that we use to manufacture and interpret research data.

#### *What We Need to Do*

I begin this final section with a simple proposition about how experimental political science can both support its own credibility and that of the discipline as a whole. The proposition is this: Make the "do-files" public. Keep lab books.

By "do-file" I mean the set of instructions that takes a scholar from a specific dataset to all of the empirical claims in a particular publication. A "do-file" contains instructions on all transformations of the specific dataset as well as a complete description of all techniques used in estimation and interpretation of this data. With a "do-file" and a common dataset in hand, replication should not be a problem.

Today, political scientists have a mixed record of keeping "do-files." Instead, what happens today is that many scholars will start with a theory and some data. They will then transform variables and try out various regression models until they get the result that fits their hunch, or perhaps a modified version of the hunch. If, for example, an initial regression does not reveal a significant effect of X on Y, they will try to square X or take the log of X, or run the regression only if X is within a certain range. With a "do-file" in hand, scholarly audiences can assess the extent to which published claims depend on such decisions. Such clarity can generate constructive debates about conceptualization and measurement.

A practical problem is that if a scholar tries different strategies to characterize a particular phenomenon (i.e., runs multiple experiments or regressions on the path from a theory to a claim), it can become hard to keep track of all the decisions that were made or why certain decisions were made. Inability to answer questions about research decisions leads to the replication problems discussed above and causes scholars to have difficulty offering credible explanations of what their results mean later on.

Lab books offer scholars a way to manage such difficulties. Researchers in fields such as chemistry use lab books to keep a record of hypotheses and observations. Best practices entail recording research activities as they occur, rather than attempting to reconstruct them from memory in later months or years. While a lab book's focal purpose is to keep members of a research team informed about current and past lab decisions, it can also be used as an essential part of a commitment to procedural transparency.

Cell biologist Jennifer Thomson (2007) describes the situation as such. "Although you may think you will remember what you did and why you did a certain experiment in a week's time, YOU WILL NOT! And nor will anyone else in your laboratory. Hence the need for laboratory notebooks. In short, a laboratory notebook is: a daily record of every experiment you do, think of doing, or plan to do; a daily record of your thoughts about each experiment and the results thereof; the basis of every paper and thesis you write; the record used by patent offices and, in the case of disputes, courts of law (in the event you file patents on your findings); a record that would enable successive scientists, working on the same project, to pick up where you left off or reproduce your results." Other sources for keeping lab books in the natural sciences include that published by the [Los Alamos National Lab](#).

While political scientists' approach to their research differs from chemists and biologists in important ways, and our needs for procedural transparency vary, the benefits of lab books to our credibility can be substantial. Table 1 sketches a framework for how political scientists can think about how to maintain a lab book. Lupia (2008) contains more information on this method for the particular case of election surveys. Chapters 6-9 of Lupia and McCubbins (1998) provide an example of the type of documentation that can be provided to scholarly audiences if detailed records are kept.

**TABLE 1. LAB BOOK PROPOSAL FOR EXPERIMENTAL POLITICAL SCIENCE**

1. *State your objectives.*
2. *State a theory.*
3. *Explain how focal hypotheses are derived from the theory if the correspondence between a focal hypothesis and a theory is not 1:1.*
4. *Explain the criteria by which data for evaluating the focal hypotheses were selected or created.*

5. *Record all steps that convert human energy and dollars into datapoints.*
6. *State the empirical model to be used for leveraging the data in the service of evaluating the focal hypothesis.*
- 6a. *All procedures for interpreting data require an explicit defense.*
- 6b. *When doing more than simply offering raw comparisons of observed differences between treatment and control groups, offer an explicit defense of why a given structural relationship between observed outcomes and experimental variables and/or set of control variables is included.*
7. *Report the findings of the initial observation.*
8. *If the findings cause a change to the theory, data, or model, explain why the changes were necessary or sufficient to generate a more reliable inference.*
9. *Do this for every subsequent observation so that lab members and other scholars can trace the path from hypothesis to data collection to analytic method to every published empirical claim.*

The first step is to state your objectives. In coordinating with others on a research program, it is useful if team members have a common understanding of the lab's objective.

Next, state a theory. The theoretical statement is an important step in demonstrating how your lab seeks to achieve the objective. The theory need not be grand. It can be something as simple as "If A, then B."

Next, make an entry about how you are deriving testable hypotheses from the theory. This step is important because many theories are capable of generating numerous hypotheses. Hence, information about hypothesis selection can be valuable to people who are attempting to understand how a particular experiment relates to a given theory. Such entries can clarify whether a particular experiment is sufficient to validate the theory, sufficient to falsify the theory, both, or neither.

Next, make an entry that describes the criteria by which data for evaluating the focal hypotheses were selected or created. Many hypotheses can be evaluated in multiple ways. Of all the ways that your hypothesis could be evaluated, why have you chosen your method? Is it the best of all possible ways to evaluate the hypothesis, or is it the best you can do given some constraint (i.e., impossible to collect new data, can only afford to run studies with student subjects, etc.)?

Next describe the means by which you are transforming

human energy and dollars into data points. These entries are perhaps the most important, particularly in circumstances where privacy concerns prevent data sharing. For example, if an explanation is meant to apply to all people, and if data creation procedures are adequately documented, then scholars have in principle the information that replication requires.

The next step involves an explanation of the estimation model used to interpret the data. For any dataset and any given hypotheses, there are often multiple empirical strategies available for evaluating the hypotheses. The lab book should include an explanation of why a particular empirical model was chosen. If the finding that this model produces is not as anticipated or suggests a revision to the theory, the need for different data, or an alternative empirical modeling specification, a subsequent entry in the lab book would explain why. Subsequent entries would then report subsequent observations and decisions as needed – writing down every estimation that is conducted, recording its attributes and, if an alternate estimation model is ultimately chosen, an argument as to why the change will generate a stronger inference.

A common source of the replication and credibility problems described above is that many scholars fail to adequately document, or make public, the steps required to produce their published claims. For example, when attempting to replicate an “If X, then Y” claim, problems occur when a scholar who originally made the claim does not adequately document how he or she manufactured variables X and Y (e.g., how he or she treated missing values).

With a lab book in hand, scholars would not be put in the embarrassing situation of not knowing how or why they got a particular result. Moreover, if they were willing to share the book with others, readers could replicate the logic that led to specification selection and the analysis itself. If lab books were honest, readers would not have to guess whether post-hoc rationalizations of observed analytic findings (e.g., stargazing) was at hand. In such cases, lab books could help readers more precisely assess the relationship between an empirical claim and a given theory.

Greater procedural transparency could lead to a more rigorous and constructive conversation about the most effective ways to draw reliable inferences about a wide range of political phenomena. It could also lead to a more constructive conversation about the conditions under which

experiments are better or worse than other methods for evaluating particular hypotheses. As the Nobel laureate and father of modern neuroscience Santiago Ramon y Cajal (1916), observed,

“What a wonderful stimulant it would be for the beginner if his instructor, instead of amazing and dismayed him with the sublimity of great past achievements, would reveal instead the origin of each scientific discovery, the series of errors and missteps that preceded it—information that, from a human perspective, is essential to an accurate explanation of the discovery.”

Above and beyond the content of the important discoveries that experimental political scientists are making, experimental political science can help to bring about a substantial advance in the discipline’s credibility by committing themselves and each other to higher levels of procedural transparency. The sooner that we as a group commit to keeping lab books and to making our “do-files” public, the sooner we can free experimental research from questions about legitimacy that affect other kinds of research in political science.

By establishing procedural transparency as part of our collective brand, experimental political science can take an important step in raising the credibility of the discipline as a whole. Were such a reputation to emerge, it is likely that other scholars would take notice and seek to improve their own legitimacy by being more honest about their procedures. Were such a sequence to occur, it could boost the credibility of the discipline as a whole by increasing confidence that published claims are derived from legitimate procedures and defensible logic and evidence.

## References

- Lupia, Arthur. 2008. Procedural Transparency and the Credibility of Election Surveys. *Electoral Studies* 27: 732-739.
- Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What they Need to Know*. New York: Cambridge University Press.
- Moravcsik, Andrew. 2010. Active Citation: A Precondition for Replicable Qualitative Research. *PS: Political Science and Politics* 43: 29-35.
- Ramon y Cajal, Santiago. 1999 (1916). Advice for a

Young Investigator. Cambridge MA: MIT Press. English translation by Neely and Larry W. Swanson.

Thomson, Jennifer A. 2007. How to Start and Keep a Laboratory Notebook: Policy and Practical Guidelines. In A. Krattiger, R.T. Mahoney, L. Nelsen, et. al., eds. Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices. MIHR: Oxford, U.K., and PIPRA: Davis, U.S.A, 763-771.

## Experimental Myths<sup>3</sup>

James N. Druckman, Payson S. Wild Professor of Political Science, Northwestern University

The existence of this inaugural newsletter reflects the increased usage of experiments in political science. The profusion of any novel methodology inevitably brings with it the risk of outpacing education about the use of the method. With this comes the potential for misperceptions - in this case, about the design, evaluation, and results of experiments. Based on my own experience from reading, reviewing, and editing journals, as well as attending conferences, I have identified six (non-exclusive and non-exhaustive) common beliefs about experimental design, all of which I believe are ill-founded. Of course many scholars do not subscribe to these "mythical" ideas; but my informal observations suggest that they reflect the beliefs of a non-trivial number of researchers. My intent is to raise attention to these important issues, rather than to delve into an extended discussion of them.

*Myth 1: The most important aspect of external validity concerns the sample of participants.*

*Redress:* External validity refers to the extent to which a study's findings generalize across populations, times, contexts, and measures. The sample is merely one of several dimensions on which one should evaluate external validity, and in many cases it is not the most relevant dimension. If the goal of the experiment is to test a causal proposition (as is typically the case), the sample will matter only if a causal moderator is posited and the sample lacks sufficient variance on that moderator.

For example, one may worry that a student sample of political novices creates problems given variation in political experience in the general population. If political experience

conditions the causal relationship under study (e.g., the effect of a persuasive message), however, this can be explored as long as experience varies within the student sample. It may be that more experienced student participants exhibit less susceptibility to the treatment (e.g., less persuadable). This implies that the average size of the causal effect is smaller in the larger population than in the student sample. It need not necessarily mean there is no causal relationship.

More generally, a problem exists only if (a) one fails to identify the feature of the sample that makes it unique or (b) there is insufficient variance on a theorized feature to test for differences in response to a treatment (see Druckman and Kam n.d.). For example, if one theorizes that a sample is problematic because it consists of highly educated individuals, then this should be directly tested by seeing whether education moderates the causal effect. A problem occurs only when the sample lacks sufficient variance in education to test for a conditional causal effect (e.g., the experimental sample consists of virtually all educated individuals). In terms of student samples, this means that one should identify the exact feature of the sample that may lead it, on average, to respond different to the treatment (compared to the population of interest). Then, if possible, test for differential reactions within the sample. While this may not be possible in a student sample if age or education level is the variable of interest, it can be done on a host of other characteristics (e.g., attitude strength, intelligence, income, political engagement).

*Myth 2: Mundane realism is more important than experimental realism.*

*Redress:* Mundane realism refers to "the extent to which events occurring in the research setting are likely to occur in the normal course of the subjects' lives, that is, in the 'real world'" (Aronson et al. 1985: 485). Experimental realism refers to whether "an experiment is realistic, if the situation is involving to the subjects, if they are forced to take it seriously, [and] if it has impact on them" (Aronson et al. 1985: 485). Without experimental realism, the causal conclusion is uncertain and thus there is nothing to generalize (regardless of the level of mundane realism). Moreover, many experiments - particularly those aimed at testing theory - are best off limiting mundane realism. Experiments should thus not be evaluated solely on the extent to which they resemble "the real world;" it is equally,

<sup>3</sup>I thank Samara Klar, Thomas Leeper, and Dustin Tingley for extremely helpful advice.

if not more, important that the experiment is designed to engage the participants to react to the stimuli as the experimenter intends.

*Myth 3: Random assignment ensures definitive causal inference.*

*Redress:* It is critical to assess the success of random assignment across conditions over all plausible moderating variables. This requires statistical comparisons on relevant variables across conditions (keeping in mind some differences will occur given a large number of conditions and variables). Acute causal inference also requires that differences between treatments are minimized so as to exclusively manipulate the precise factor under study (e.g., studies on the impact of racial visuals should use the same visuals that vary only by skin color).

*Myth 4: Experimental control means we understand the treatments.*

*Redress:* An experimentalist should not assume that his or her perception of a treatment stimulus is the same as that of the experimental participants. It is important to conduct pre-tests and manipulation checks to ensure that experimental participants perceive stimuli as the experimenter intends (e.g., an experiment on reactions to corruption should not assume that participants view corruption in the same way as the experimenter; see, e.g., Redlawsk and McCann 2005).

*Myth 5: Testing mediators and moderators is straightforward in experimental settings.*

*Redress:* Exploring moderators needs to be theoretically guided such that the focus is on hypothesized subgroups (Kent and Hayward 2007). That is, one should not engage in evaluating interactions between the treatment and any possible demographic variable. By chance alone, one would discover a significant moderator if the search covers a large number of demographic and/or political variables. Instead, experimentalists need to theorize on what individual features will condition the causal relationship and test for those specifically. Experimentalists also should consider, when possible, blocking on theorized (and non-manipulable) moderators (see Levin 1999 and Moore, this newsletter).

Mediation analyses - that seek to establish the path through which a treatment variable influences a response variable - are nearly always imperfect since omitted variable bias occurs (see Bullock and Ha n.d.), though formal

methods for sensitivity analyses are now available (Imai et al. n.d.). The ideal way to explore mediation is to conduct a distinct experiment on mediation, though the selection of the right design should be done with great care (Imai et al. 2010).

*Myth 6: Reporting standards of experiments is the same as reporting standards for surveys.*

*Redress:* Political scientists are just coming to grips with the standards for reporting surveys, as laid out by the American Association for Public Opinion Researchers (see Hillygus 2009). Unfortunately, these standards do not ask for the reporting of information critical to experimentation, such as randomization checks, manipulation checks, pre-test results, and so on (see Gerber, Doherty, and Dowling 2009, Lupia, this newsletter). Going forward, researchers should work actively to establish thorough and uniform standards for reporting experiments. This not only is critical for transparency but also to facilitate replication and meta-analysis.

In their highly influential treatment of experiments, Campbell and Stanley (1963: 3) explained that we must “justify experimentation on more pessimistic grounds - not as a panacea, but rather as the only available route to cumulative progress. We must instill in our students the expectation of tedium and disappointment and the duty of thorough persistence. We must expand our students’ vow of poverty to include not only the willingness to accept poverty of finances, but also a poverty of experimental results.” While this may be an overstatement, a bit of caution is in order and that is what I hope to have provided here. For in order to advance science, experiments need to be theoretically motivated, carefully constructed, and skillfully analyzed. The alternative is misuse and, likely, an ill-founded illusion of progress.

## References

- Aronson, Elliot, Marilyn B. Brewer, and J. Merrill Carlsmith. 1985. “Experimentation in Social Psychology.” In Gardner Lindzey, and Elliot Aronson, eds., *Handbook of Social Psychology*, 3rd Edition. New York: Random House.
- Bullock, John, and Shang Ha. N.d. “Mediation Analysis Is Harder than It Looks.” In James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds., *Cambridge Handbook of Experimental Political Science*. New York: Cambridge. Forthcoming.

Druckman, James N., and Cindy D. Kam. N.d. "Students as Experimental Participants: A Defense of the "Narrow Data Base." In James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds., *Cambridge Handbook of Experimental Political Science*. New York: Cambridge. Forthcoming.

Gerber, Alan S., David Doherty, and Conor Dowling. 2009. "Developing a Checklist for Reporting the Design and Results of Social Science Experiments." Presented at Experiments in Governance and Politics Network meeting, Yale University, April 24-25, 2009.

Hillygus, D.S. 2009. "The Need for Survey Reporting Standards in Political Science." In Gary King, Norman Nie, and Kay Schlozman, eds., *The Future of Political Science: 100 Perspectives*. New York: Routledge.

Imai, Kosuke, Luke Keele, and Dustin Tingley. N.d. "A General Approach to Causal Mediation Analysis." *Psychological Methods*. Forthcoming.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2010. "Experimental Designs for Identifying Causal Mechanisms." working paper, Princeton and Harvard University.

Kent, David, and Rodney Hayward. 2007. "When Averages Hide Individual Differences in Clinical Trials." *American Scientist* 95: 60-65.

Levin, Irwin P. 1999. *Relating Statistics and Experimental Design*. Thousand Oaks, CA: Sage Publications.

Redlawsk, David, and James McCann. 2005. "Popular Interpretations of Corruption and their Partisan Consequences." *Political Behavior* 27: 261-284.

## Innovations in Experimental Design: Lessons from "Get Out The (Free And Fair) Vote" in Africa<sup>4</sup>

Kristin G Michelitch, NYU Ph.D. Candidate,  
[kgm254@nyu.edu](mailto:kgm254@nyu.edu)

**Violence. Vote-buying. Ballot fraud.** The question isn't so much how to get people to turn out to vote in sub-Saharan Africa. The question is more how to get citizens to turn out of their own free will, cast votes for parties that haven't just offered wads of cash for them, and be confident that their ballots weren't destroyed in an effort to manufacture a more \*favorable\* result.<sup>5</sup> This review highlights a set of field experimental interventions aimed at getting out the free and fair vote in three recent elections in sub-Saharan Africa. Not only does this research represent a huge break into large and substantively uncharted waters, but we can also learn from some clever design and measurement techniques employed by the authors.

- Vicente, Pedro (2008) *Is Vote Buying Effective? Evidence from a Field Experiment in West Africa* Working Paper (in Sao Tome & Principe)
- Collier, Paul and Pedro Vicente (2010) *Votes and Violence: Evidence from a Field Experiment in Nigeria* Working Paper
- Aker, Jenny, Paul Collier, and Pedro Vicente (2010) *Is Information Power? Using Cell Phones during an Election in Mozambique* Working Paper

Interventions were the following (please click on hyperlinks to see visual, audio, and video props). The authors undertook three **campaigns** during the 2009 Mozambique elections. By utilizing new communication technologies in a number of novel ways, the authors were able to cover a whopping 44% of the Mozambique population. First, the relatively straightforward voter education intervention involved distributing **leaflets** regarding when and how to vote along with verbal explanation and discussion. Subsequently, citizens were sent

<sup>4</sup>This graduate student article was "commissioned" by the editor to start a recurring long form section of the newsletter where a graduate student has the opportunity to discuss new work in their field. Future submissions of this nature are highly encouraged.

<sup>5</sup>Africa as a whole has the lowest percent of voting individuals out of those individuals eligible to vote compared to other world regions (naturally this masks large intraregional variation). See **IDEA**.

10 SMS daily starting two weeks prior to the election about the date, parties, candidates, and the like. The idea was to increase awareness of voting and election procedures to increase turnout. Second, in the fraud/violence hotline intervention, leaflets (along with verbal explanation/discussion) were distributed throughout the locality about the availability of a nation-wide hotline to be texted if someone saw a incident of fraud or violence. Local correspondents in each locality verified the report and if it was true, the media was given the report and the report was shared with all the respondents via SMS. Two weeks before the election, respondents were sent reminders about the existence of the hotline by SMS. The actual SMSs were publicly available on the internet. This project utilized an **ushahidi platform**, which allows anyone to gather distributed data via SMS, twitter, email or web and visualize it on a map or timeline. The goal was to increase confidence that fraud was being monitored and moreover, monitored by voters themselves. A combination of the voter education with the fraud/violence hotline campaign was the third intervention. This intervention was an interaction of the voter education and hotline treatment content but the medium of delivery was different. Instead of door-to-door, the content of the messages was delivered through a weekly **newspaper** subscription.<sup>6</sup>

In the 2006 Sao Tome & Principe (STP) election the author waged a **campaign** against the ubiquitous vote-for-cash (or other gifts) exchange between party representatives and voters. The campaign emphasized that vote-buying is illegal and urged that citizens should “vote their conscience” regardless of whether they took the money. The medium of communication was door-to-door personal distribution of **leaflets** and discussion with citizens. In the Nigerian 2007 election, the authors wanted to combat violence and threats used by party thugs to intimidate voters into voting for a particular candidate or not at all. An **intervention** was conducted to raise awareness that the use of political violence at election time is wrong and urged

citizens not to vote for politicians who use violence. The content was delivered via town meetings, **popular theater**, **radio jingles** and the distribution of campaign material such as **hijabs**, stickers, and tshirts. The idea was that by waging community wide campaigns, citizens would be empowered to resist intimidation knowing that general resistance to intimidation amongst the citizenry had increased. Anticipating growing resistance, parties would then decrease the use of intimidation as a strategy to win votes.<sup>7</sup>

Procedures are similar across the studies. The authors conduct the first wave of a panel survey in both intervention and baseline areas. Within the survey, enumerators administered the first dose of the intervention directly to subjects in intervention localities before collecting data on dependent variables. The interventions were also administered throughout the intervention localities to non-respondents (randomly in the case of door-to-door leaflet distribution, but non-randomly in the case of popular theater). Many interventions continued for a number of weeks up to the election. After the election, the authors conduct a second wave of the panel survey. The authors then employ difference in difference (DID) estimation of various causal effects.

To concisely communicate the treatment effects across these studies, Table 2 shows the treatments in the first row and a selection of dependent variables in the first column. Inside the matrix are the treatment effect directions.<sup>8</sup> The interventions were largely effective in doing what they aimed to do: turning people out to vote, reducing violence and perceptions of violence, and reducing perceptions of the effectiveness of vote buying. Yet some interventions produced somewhat undesirable results. The anti-fraud/violence hotline campaign increased the perception of violence and decreased the perception of free and fairness in the election. Most treatments benefited the incumbent party in a world where the incumbent already has a huge advantage. The papers serve as nice examples of how authors should report results from multiple casual

<sup>6</sup>Citizens had never received a newspaper before.

<sup>7</sup>The authors argue that opposition to intimidation is a collective action problem in that it is individually costly to resist, but the more people resist, the lower the cost to each individual of resistance. In my opinion, the effectiveness of this manipulation relies not on urging an individual not to vote for violence-using politicians, but informing the individual that yet other individuals had been successfully convinced to resist violence. To truly evaluate the theory, a future campaign could focus on changing individuals' beliefs about the distribution of willingness to resist in the population in order to change the target individual's own willingness to resist (thereby impacting turnout). Perhaps an ushahidi platform could be utilized as in the Mozambique study.

<sup>8</sup>Note that these studies are in working paper form so the authors may produce yet other results. Note that question wording and number of questions used to measure each dependent variable changes in each study. Please see the studies themselves for the various estimations of causal effects.

effects estimation procedures and robustness checks. While I will leave the details on the results to the reader's own perusal of the papers, I would like to highlight and discuss just a few novelties in the authors' design.

The authors attempt to reduce measurement error due to untruthful responses in the dependent variables in a variety of ways that may be useful to other experimenters. First, one can imagine that if a subject undergoes, for example, an anti-violence campaign, that he/she may feel obligated to say that violence has decreased even if it has not. The authors are worried that such a "response conformity bias" will inflate the magnitude of the DID estimate. Let me introduce a bit of notation to explain the problem and proposed solution in simple terms. Suppose we have the mean of the outcome variable  $\bar{y}_M^t$  where  $t$  is 0 and 1 for the first and second survey wave respectively and  $M$  is the exposure to the manipulation with support  $[0, m]$ . Assume for simplicity that the treatment effect is positive and there exists a conformity bias function  $c(M)$  constant across individuals. Usually, experimenters undertaking a DID approach choose to undertake the initial baseline survey and measure the dependent variables in all localities *before* administering the intervention, yielding Equation 1. The conformity bias terms  $c(0)$  drop out when subtracting the overtime differences in the baseline group. Bias remains  $c(m) - c(0)$  when subtracting the overtime difference in the treatment group, inflating the DID estimate.

$$\hat{\beta}_{DID} = [(\bar{y}_{M=m}^{t=1} + c(m)) - (\bar{y}_{M=\epsilon}^{t=0} + c(0))] - [\bar{y}_{M=0}^{t=1} - (\bar{y}_{M=0}^{t=0})] \quad (1)$$

$$\hat{\beta}_{DID} = [(\bar{y}_{M=m}^{t=1} + c(m)) - (\bar{y}_{M=\epsilon}^{t=0} + c(\epsilon))] - [\bar{y}_{M=0}^{t=1} - \bar{y}_{M=0}^{t=0}] \quad (2)$$

By administering a small dose  $\epsilon$  of the intervention, *before* collecting data on the dependent variables in the first survey wave in intervention localities, the authors claim to control for conformity bias. This design yields Equation 2. In order for the bias to wash out, the authors (implicitly) assume that  $c(\epsilon) = c(m)$ , that is, that  $c()$  is discontinuous and no matter what the dosage of the intervention, it kicks in the same degree of conformity bias. However, if the bias is

continuously increasing in the manipulation, meaning that the more anti-violence campaigning, the more likely respondents will feel compelled to say there was an effect on violence, then the bias is  $c(m) - c(\epsilon)$  and we have (only) gotten a reduction in the bias of  $c(\epsilon) - c(0)$ . This design decision may constitute a vast improvement over a typical survey experimental design of a single survey wave if the researcher a priori surmises the conformity bias to be large.<sup>9</sup> Yet even if conformity bias is a small worry, the design may provide other benefits. Given randomization has produced balance across treatment groups, we could estimate the effect of a one-shot versus a large dose of the treatment. These points are underexplored and undersold in the papers.

The second type of design feature used to reduce measurement error is by collecting clever behavioral measures of key dependent variables. It is well established that respondents overreport turnout and otherwise self-censor the truth if it conflicts with social desirability.<sup>10</sup> One measure of turnout is whether the respondent showed his/her inked finger<sup>11</sup> without hesitation when asked to do so by the enumerator. A second was based on a composite score of a battery of questions about the election-day experience (e.g. with whom the person went to the polls, what time, how long they waited, how long the line was). The enumerator collected this data but for each item also coded if it was convincing that it actually happened. A third measure was based on another battery of questions which focused on knowledge of ballot station facts (e.g. number of ballot papers, whether there were photos of the candidate, number of ballot boxes, whether they were colored). The last measure was the enumerator's general belief about whether the respondent turned out given the interaction with the subject over the entire part of the survey about turnout. More studies should similarly measure engagement in an activity by asking respondents questions that only someone who has actually engaged in an activity would be able to answer. One caution in the Mozambique study, however, is that in the case of turnout, some voter education interventions in the current authors' studies describe the balloting process and may enable non-voters to look like

<sup>9</sup>The bias is easy to see in this case:  $(\bar{y}_{M=m} + c(M)) - (\bar{y}_{M=0} + c(0))$ .

<sup>10</sup>Note that, while other studies may consider a comparison of survey to aggregate data, we can be certain that the numbers reported by the National Electoral Commissions in sub-Saharan Africa are fudged (often over 100% turnout rates are officially reported).

<sup>11</sup>In many countries, voters must dip a particular finger in a vat of indelible ink so they may not vote a second time. This ink is supposed to linger for between 72 hours to 3 weeks.

voters - actually increasing measurement error and falsely inflating treatment effects. Other studies may also benefit from the technique of leveraging the enumerator's interaction with the subject.<sup>12</sup>

The studies further tried to capture behavioral measures of other key intervention effects. For example, respondents could choose to send a postcard that the media should raise awareness in undermining political violence in the Nigeria study. Because sending the postcard is costly, it reveals that the citizen is at least willing to stand up against violent intimidation through costly action. In the Mozambique study, voters are given the opportunity to SMS (at a cost) a set of policy preferences to the president elect via a newspaper. The authors argue that this behavior reveals demand for accountability better than expressions. Unfortunately, noting that there is only a treatment effect on this behavior in the newspaper intervention may alternatively indicate an increase in the ability to have policy preferences, trusting the newspaper to display the result, or the novelty of seeing your contribution in the new newspaper you've been getting.<sup>13</sup>

The authors have to grapple with spillover in their studies and design two major approaches to estimate the degree of spillover effects. First, geographic distance to treatment areas was utilized to capture treatment exposure amongst those in baseline areas. Second, the authors interview an additional set of individuals in intervention localities during the post-election survey wave to assess the effect of receiving the intervention "indirectly" (not "directly" by being a panel survey respondent). However, this particular design feature allows a modest exploration of spillover but more importantly gives us a clue about (1) the role of personal contact in the delivery of the interventions and (2) whether heterogeneous treatment effects exist across those forced to get the treatment versus allowed to select in or out.

First, the American GOTV literature has generally found that increasing levels of personal contact with another human in the message delivery increases the effectiveness of turnout campaigns. Especially in Africa, personal contact and feeling flattered because someone from outside cares to hear your political views may amplify the intervention's

effect versus receiving the intervention in the absence of sustained contact and direct treatment from a survey enumerator. For example, being directly approached by the antiviolence campaign, given a "no to violence" hijab and personally invited to popular theater may have a wildly different effect than seeing other community members with the hijab and stopping by the theater on your way home.

Second, if the authors have data on treatment exposure (e.g. how many times did you see the roadside show?), there may also be room to explore whether heterogeneous causal effects exist conditional on whether individuals were forced to take the treatment (the panel respondents in intervention areas) or were allowed to self-select into or out of the treatment (those who were interviewed only in the post-election wave in intervention areas). For example, some individuals in STP thought vote buying was bad and others good.<sup>14</sup> Treatment effect may be conditioned on this opinion. One could compare treatment effects conditional on opinion of vote buying for panel survey respondents who were forced to take the anti-vote buying treatment, to those in just the post-election wave who could endogenously self select in or out of the treatment. In these large campaigns where researchers have lack of control, it would be a huge contribution to measure treatment exposure and exogenous versus endogenous selection.

Because these interventions broach some serious topics - violence, vote-buying, and fraud - on which we've never (or barely) experimented before, it throws up some meaningful questions about general equilibrium effects. Here parties may change strategy in response to such large-scale interventions in the electorate. Parties have a set of strategies to target various constituents, which all probably have increasing marginal costs. By undertaking an intervention meant to shift around the cost curve of one of their strategies, which other strategy, now made relatively cheaper, will be utilized? Perhaps worryingly, if you do an anti-vote buying intervention, will parties switch to violence unless you do an anti-violence intervention as well? Parties may also work around the intervention. For example, it is well known that parties often take the indirect route to votes

<sup>12</sup>For example, income is notoriously hard to measure in Africa, but some authors have started utilizing the enumerator's size-up of the respondent's living conditions to tap into income. Spending 1 hour interviewing the respondent in his/her home is quite telling.

<sup>13</sup>One could even imagine this last task - allowing voters to communicate with politicians about their preferences - as a potential future treatment that might affect turnout or other forms of political participation.

<sup>14</sup>It's the only time they get something from the politicians!

by offering large sums to tribal chiefs, religious leaders, or “Big Men” to influence their constituents to turn out and vote for a certain party. Did parties switch to doing that rather than directly vote-buying from citizens in the STP experiments? When parties catch wind of the interventions - which they must in these studies by partnering with the National Electoral Commission - will they double their cheating efforts in localities outside of the surveyed area? In the future, it may benefit studies of this kind not only to hire local correspondents in the surveyed areas to document such things as violence or fraud, but also hire them to collect data on party strategies generally and employ them in yet other areas in and out of the survey area.

Not only might the interventions affect party strategy, but it might do so differentially based on incumbent or opposition status. Looking at the Table ??, if anyone benefited, it was the incumbent. First, one concern is that not only the message, but the message bearer matters. In particular, by displaying a partnership with the National Electoral Commission in all of these interventions, citizens may have thought the message was coming from the incumbent party.<sup>15</sup> Differential effects for party behavior could also derive from differential intervention effects on opposition versus incumbent voters. Thus, one of the biggest improvements the authors could make in the papers is to explore the possibility of heterogeneous treatment effects based on partisanship at the individual-level, locality-level,

and both interacted. For example, it is easy to see that anti-violence campaigning may motivate incumbent voters, while demotivating opposition voters to turn out if violence is the only tool the opposition party has to win. The authors should also demonstrate whether there is balance in aggregate partisanship across baseline and treatment localities, as lack of balance of this very important covariate may drive results, and has a high probability of occurring.

Because we are just at the beginning of this research agenda, interested researchers also face the same challenges the US American GOTV research faced - disaggregating campaigns into smaller parts to understand exactly what was most effective. The campaigns in these studies were each a bundle of treatments together - which part of the intervention was driving the results, or was it a synergy between multiple elements? The authors could also tell us which ones are the most cost effective. Another question is whether interventions will have the same effects after repeat doses. The marginal effect of getting a newspaper subscription going from no newspaper to one newspaper may be wholeheartedly reduced 10 newspapers later. The novelty of receiving SMS may wear off and voters may get annoyed receiving a 10 texts per day for 14 days. On the contrary, some interventions like the hotline become more effective when voters learn how to use it better. Hark the christening of the GOT(FAF)V research agenda...

---

<sup>15</sup>The National Electoral Commission is not non-partisan, but fully under the control of the incumbent.

Table 1: Table 2: Interventions, Dependent Variables, and ITT/ATE by Country

Intervention ⇒ DV ↓	STP 2006 Anti-Vote Buying	Nigeria 2007 Anti-Violence	Mozambique 2009 How To Vote	Mozambique 2009 Fraud/Violence Hotline	Mozambique 2009 Newspaper
Perceptions of Vote-Buying Being Effective	↓				↑
Frequency of Vote Buying	(↓)				
Taking the money and voting your conscience	↑				
Perception of Violence		↓		↑	↓
Perception of Citizen Tolerance for Violence		↑			
Violence Intensity		↓			
Sent "I am behind campaign" Postcard		↑			
Perception of Free/Fair Election		↑	(↑)	↓	
Sent SMS about policy preferences					↑
Turnout	(↓)	↑	↑	↑	↑
Interest in Elections/Political Knowledge			↑	↑	↑
Incumbent Voteshare/Liking Incumbent	↑	↑*		↑	↑

\* Opposition voters switching to absenteeism and absentees switching to be incumbent voters. Arrows in parentheses indicated limited evidence. Target population: eligible voters. Samples: The authors start by getting a representative sample of localities, randomly select a subset of the localities for intervention(s) and baseline, and randomly select respondents within the localities using standard nth household survey design.

## Section News and Announcements

We are pleased to announce the APSA Experimental Section Officers

1. President: Don Green
2. President-elect: Jamie Druckman (2011)
3. At large council: Rose McDermott (2010), Mike Tomz (2010-2011), Lynn Vavreck (2010-2011)
4. Treasurer: Kevin Esterling (2010-2011)
5. Secretary: Costas Panagopoulos (2010-2011)
6. Newsletter editor: Dustin Tingley (2010-2011, renewable and appointed by President)

We are pleased to announce awards committees. Awards are made possible through joining the APSA section.

1. Best Dissertation in prior calendar year: Sean Gailmard (chair), Bethany Albertson, Nick Valentino, Shana Gadarian
2. Best Paper at prior APSA: Josh Tucker (Chair), Rose McDermott, James Gibson, Eric Dickson
3. Best Book: Ted Brader (chair), Susan Hyde, Maccartan Humphreys, Ismail White

## Upcoming Events

- Fourth Annual NYU-CESS Conference on Experimental Political Science call for papers:

We are pleased to announce the Fourth Annual NYU-CESS (New York University Center for Experimental Social Sciences) Conference on Experimental Political Science on Friday, March 4th, 2011 and Saturday, March 5th, 2011. The Conference is an annual event that brings together researchers interested in experimental methodology in political science broadly. That is, we welcome the participation of scholars who work in the field and those who work in the lab as well as the participation of political psychologists and political economists. Furthermore, we welcome the participation of scholars who are not experimentalists themselves but are interested in learning and discussing experimental methods as well as those interested in the relationship between experimental methods and analyzing observational data in political science. We invite submissions of papers for possible presentation at the fourth annual conference. Paper presenters from outside the NYC area will be offered a small stipend to help defray the expenses of attending the conference. If you would like to submit a paper, please go to <http://cess.nyu.edu/CallForPapers/submit/612e18f76cfbf89a32588a340912e4d7> The deadline for submissions is November 15, 2010. Authors notified of decisions by December 15, 2010.

- Southern Political Science Association Mini-Conference on Experimental Political Science

January 5-8, 2011, New Orleans, Sponsored by the APSA Experimental Research Organized Section this mini-conference will bring together papers by Rebecca Morton and Jean-Robert Tyran, Kevin Arceneaux and Martin Johnson, William Minozzi and Jonathan Woon, Christopher Mann, Rick Wilson and Catherine Eckel, Dan Myers and Dustin Tingley, Ngoc Phan, Cindy Rugeley and Gregg Murray, Laura Paler, Noam Lupu, and Brad LeVeck. The conferences promises to be exciting!

- Call for participation: NY Area Graduate Student Experimental Working Group

The NY Area Graduate Student Experimental Working Group invites graduate students from the New York City Area to participate in the upcoming workshops of 2010-2011. Dan Myers, Kristin Michelitch and Dustin Tingley founded this group in 2009, hoping to create a forum for constructive feedback during the critical design stage of experimental research. Experimental designs often do not come under sufficient scrutiny before experiments are conducted. Further, we recognized that our work could benefit from exchange across the political economic and political psychological persuasions of experimentation. At workshop meetings, we discuss 3-4 papers for 1 hour apiece. Each submission is assigned a discussant from a different university and participants are expected to have read all the papers and come prepared to discuss. Upcoming meetings are from 1pm-5pm: November 12th 2010 at Princeton, February 4th, 2011 at NYU, and April 29th, 2011 at Location TBA For further information, to submit a proposal to present, or to be added to the mailing list for this group, please visit <http://www.princeton.edu/cdmyers/NYworkinggroup.htm> or email [cdmyers@princeton.edu](mailto:cdmyers@princeton.edu) or [kristin.michelitch@nyu.edu](mailto:kristin.michelitch@nyu.edu)

- IPSA Summer Experimental Class

The International Political Science Association is sponsoring a methods summer school for MA and PhD candidates from January 31-February 11, 2011 at the University of Sao Paulo, Brazil. Rebecca Morton will be teaching an Experimental Political Science Course as part of the summer school. For information on the course see:

<http://summerschool.ipsa.org/mission>

- Fourth Annual West Coast Experiments Conference call for papers:

The Fourth Annual West Coast Experiments Conference will be held at Caltech, in Pasadena, CA, on May 6, 2011. This will be an all day conference, and will bring together researchers interested in advances in experimental methods for political science. It is sponsored in part by the Caltech-USC Center for the Study of Law and Politics and in part by the Caltech Division of Humanities and Social Sciences. This year's co-organizers are Kevin Esterling, Jonathan Katz, Mathew McCubbins, and Nicholas Weller. There are no registration fees for the west coast conference, and all meals for the day will be provided to all registered attendees. Space will be limited. Nominations for paper presentations can be sent to [kevin.esterling@ucr.edu](mailto:kevin.esterling@ucr.edu). More details to come.

[Back to Contents](#)

## Famous Experimentalist Quotes

*"I'm experimental by nature... always exploring my creativity."* — Christina Aguilera