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.

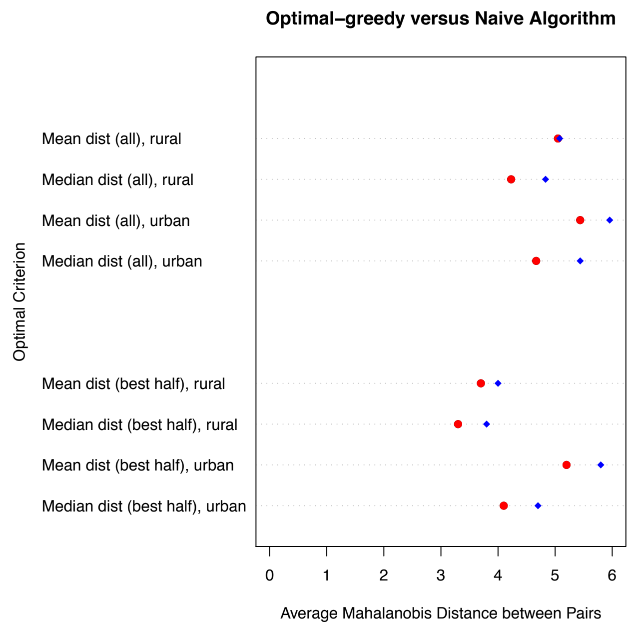
To implement blocking in an actual experiment, the first decision is to choose the variables to block on. You can block on a large set of covariates, including discrete and continuous measures. Blocking should focus on variables likely to affect the outcome of interest. Similarly, for any important subgroup 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.[1] 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. 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.[2]



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

[1] 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

[2] The eight comparisons represent two types of units (urban/rural), two subsets of units (all/best half), and two global meaures of optimality (mean/median distance).

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