The Experimental Political Scientist

In this issue

From the Editor

Welcome to the next issue of the Experimental Political Scientist. Leading off is a discussion by Ryan and Broockman on how to use Facebook for field experiments. Next Redlawsk takes us through the history and innovations of the Dynamic Process Tracing platform. Kramon and Weghorst deliver a fascinating analysis of asking sensitive survey questions using list experiments in developing countries. Finally, we have two excellent book reviews. Jonathan Woon takes on the controversial book by Clarke and Primo and Anand reviews the pathbreaking book "The Social Citizen" by Sinclair. Thanks to all the contributors. Special thanks to the "field" editors: Dan Myers, Jonathan Renshon, and Kristin Michelitch. PLEASE send contributions (or ideas) for the spring newsletter, deadline April 15th. Happy experimentation!

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. Previous issues available. Send submissions to the editor.

In the discipline

Sadly Hurricane Irene disrupted APSA. There was lots of great stuff on tap due to the hard work by conference section chair Ken Williams (who has a new game theory book out, Introduction to Game Theory: A Behavioral Approach!). We'll look forward to the upcoming experimental conferences (NYU and West Coast), and everyone can expect to see some great experimental papers at Midwest and other conferences. APSA has approved our journal and Rebecca Morton and Joshua Tucker will soon be circulating a call for papers. The journal will be an exciting place for path breaking experimental work across the sub-fields. Stay tuned for this and other announcements from our Section leadership.

The May 2013 issue will be my last as editor, I'm hoping to get some really great submissions! I particularly invite submissions on best practice in teaching experimental methods but also on the ways experiments are being brought into the classroom.

Dustin Tingley, Editor
Harvard University
Government Department

Facebook: A New Frontier for Field Experiments

Timothy Ryan and David E. Broockman
University of Michigan and University of California, Berkeley
TJRyan@umich.edu
broockman@berkeley.edu

On September 14th, 2012, someone (we know not who) became the one-billionth member of the online social network Facebook (Facebook 2012a). As this figure suggests, online networks in general - and Facebook in particular - have rapidly become central to American social life. More than half of the United States' approximately 154 million Facebook users visit the website every day (Hampton et al. 2011); and for many of these users, Facebook has become the default platform to share photos, communicate with friends near and far, and announce life developments both mundane and profound. The social network is also becoming ubiquitous in new ways: more than half of Facebook's users now connect to the site with mobile devices, meaning that millions of individuals carry the network's influence with them as they go about their lives (Facebook 2012a).

Just as Facebook is becoming ever more central to American social life, the network is also playing a growing role in the nation's politics. Two-thirds of Americans on social networks use them to engage in political activity such as sharing thoughts about issues and following the activities of candidates (Raine et al. 2012).[1] For their part, political elites also see social media as an important tool for communicating with the public. At least four fifths of members of Congress now have official Facebook pages, and many maintain a separate page for campaign purposes (Congressional Management Foundation 2011a). Majorities of congressional staffers likewise report using social media to gauge their constituents' views and to spread messages to people they typically have difficulty reaching (Congressional Management Foundation 2011b). Mitt Romney's Digital Director Zac Moffat recently echoed this perception, remarking, "[Y]ou can't reach everyone by just television alone anymore. . . . If you don't try and have conversations [over social media], you will just miss people" (quoted in Simon 2012). See Figure 1 for an example of some of these efforts in the form of politicians' Facebook advertising.

Students of American politics have two compelling reasons to take an interest in social media in general, and Facebook in particular. First, Facebook's ever-expanding size and reach clearly make it of interest in and of itself as a new political phenomenon. How will the new community, with its new ways to follow and participate in politics, impact the broader political system? Second, the technology that Facebook incorporates gives researchers new leverage on classic questions about political behavior: what makes people attend to politics, remember political information, and participate in political conversations? We expand on both these themes below, giving particular attention to the possibility of using Facebook as a new environment for conducting field experiments.

A Distinctive Political Community

---

[1]Indeed, such political activity is so common on Facebook that enterprising software engineers have even developed applications that allow users to block their friends' apparently irksome political activity from view.

As a forum for political interactions, Facebook is unique in several important ways. Consider, first, the staggering volume of potentially analyzable data. Users log 3.2 billion "likes" and comments on the site each day (Facebook 2012b), all of which are catalogued by Facebook's software. Of course, the primary reason this information is recorded (other than for users to see) is for marketing purposes. As a byproduct of these efforts, however, the website is generating a vast chronicle of opinion - what people around the world find interesting, motivating, attention grabbing, objectionable - that has no historical analogue.

Aside from being plentiful, the data have special characteristics. First, the typical Facebook user pays scant attention to the fact that nearly every click and keystroke could be analyzed for research. In light of perennial concerns that self-reports about political participation and information exposure are systematically biased and that individuals change their behavior when they are aware of being studied, this high level of naturalism could address measurement problems that have long vexed the study of political behavior (e.g., Sears and Freedman 1967).

Facebook also presents researchers and practitioners with a fresh opportunity to focus an analytical spotlight on the young. As most readers will be aware, the attitudes and behavior of younger Americans have become increasingly difficult to measure as they forgo (in disproportionate numbers) the land-based telephone lines that survey researchers habitually target (Blumberg and Luke 2010). Researchers also might need to give separate attention to the information streams that reach young citizens, as young people are much less likely to attend to traditional forms of media, such as television (Chozick 2011). Facebook's demographic makeup is skewed in a way that countervails this deficiency, with about 40% of its U.S. users under the age of 30 (see Table 1).

Conducting Experiments on Facebook

With all the political activity being logged on Facebook, the network seems ripe for study. A particularly enticing possibility is that Facebook might serve as an engine to continue the estimable tradition of leveraging the power of random assignment to study political communication (e.g., Hovland, Lumsdaine, and Sheffield 1949).

With so much activity being recorded, and with advertising space readily available to purchase, it would seem that randomized trials would be straightforward to conduct. Our explorations into the possibility find that it is indeed possible to conduct randomized trials on Facebook, but doing so requires the navigation of some obstacles.

The biggest complication concerns the mechanism of random assignment. Of course, perhaps the most intuitive way to assign any treatment is the individual level. This is the approach used by Bond and colleagues in their recent study of voter turnout messages on Facebook. On Election Day in 2010, each U.S. Facebook user who logged in to his or her account was randomly assigned to one of three conditions: viewing a large banner at the top of the site with a "social message" that encouraged voting and showed which friends had voted already, an "informational message" that was largely the same but omitted the details about friends' behavior, or no message at all (Bond et al. 2012).

Though such individual-level assignment is typically ideal, this avenue is not open to most researchers: Bond and colleagues closely collaborated with Facebook researchers and programmed a

highly visible intervention into Facebook's own core infrastructure.[2]

Happily, implementing well-powered field experiments on Facebook does not require randomizing at the individual user level. Rather, experimenters can apply a time-honored field experimental technique: randomly assigning clusters of individuals (cf. Arceneaux 2005). Many experiments on education policy randomize at the level of schools or classrooms, rather than students; get-out-the-vote canvasses are typically randomized at the level of households or neighborhoods, not voters; and television and radio ads must be randomized at the level of media markets, not individual devices (e.g. Panagopoulos and Green 2008). With some creativity, Facebook's advertising interface permits a clustered randomization of this same kind.

The approach we devised makes use of Facebook's ability to target ads at particular user demographics. Using Facebook's interface, one can limit who will see an ad based on traits that nearly all members report in their profiles, the chief ones being age, gender, and geographic location (country, state, city, or zip code). Furthermore, Facebook allows users to easily purchase many (thousands) of ads and target each separately. It is thus possible for a researcher to "slice up" a population into many clusters defined along demographic characteristics, randomly assigning each cluster to a particular experimental condition.

Our preferred approach for doing so is to use mutually exclusive characteristics to create disjoint clusters, ensuring no crossover with respect to which users could be exposed to a particular ad (see Table 2). For instance, in one study, 23-year-old males living in San Francisco might be assigned to view one ad, but 23-year-old females in the city another. With several targeting dimensions available to distinguish users, and with some dimensions taking on a large number of discrete values (age and geography in particular), it is quite easy to generate a large number of such clusters even among a relatively small population. For example, the San Francisco Bay Area alone contains 101 municipalities. Combined with 53 discrete, targetable ages (13 year olds, 14 year olds, 63 year olds, 64 year olds, 65+ year olds) and two genders, its Facebook users could be cleanly divided into over 10,000 clusters based on these categories alone. Randomizing the delivery of some treatment at the level of these ¿10,000 clusters would yield an experiment less well-powered than randomizing at the level of all 7 million inhabitants of the area, but still one that could boast ample statistical power for most applications.[3]

It is also possible to target advertisements on the basis of affiliations and interests members list in their profiles (e.g., "Democratic Party" or "sailing"). Though these targeting criteria are less attractive for the purposes of randomization (as they are not mutually exclusive in the way gender and age are), researchers interested in investigating how particular characteristics moderate treatment effects could use them to compare treatment effects among different populations.

Aside from lower statistical power, another limitation of randomizing advertisements is that, because

---

[2]See also Grimmer et al. (2012), who randomize the delivery of messages at the individual level with the help of a user-installed application, though thereby alert participants that their Facebook accounts are under study. Some researchers also use Facebook advertisements as a tool for subject recruitment (see Samuels and Zucco [2012] for further discussion).

[3]One issue with this process is that individuals may log into the site in multiple locations, so in practice researchers should take care to use geographies that individuals are unlikely to cross in the course of their daily lives.

these treatments are woven into Facebook's existing architecture, they preclude such novel and potent interventions as Bond et al. employ. We see this limitation as very slight. Field experimenters often (and rightly) point to a high level of realism as a chief virtue of their approach. The limitation thus amounts to little more than a commitment by researchers to work within the same framework - its tools and constraints - that political actors themselves do as we examine the implications of their behavior.

Measuring Outcomes

Having conducted clustered random assignment of advertisements, there are a number of possible dependent measures that might be of interest. One class of measures concerns individuals' activity on the social network itself, many features of which Facebook reports to advertisers. For instance, one of us recently reported the results of an experiment that exposed Facebook users to emotionally evocative political advertisements, finding anger-inducing ads (relative to anxiety-inducing and emotionally neutral ads) to double the likelihood that Internet users would click the ad to learn more information on a political topic (Ryan 2012). For this study, the click-through percentage in each cluster served as a naturalistic measure of information seeking.

The set of questions for which clicks would be a sufficient dependent variable is large - we are conducting further studies of partisan selective exposure and the effect of negativity - but there are also opportunities for more sophisticated designs. For instance, Facebook allows users to create individualized pages, events, and applications, each of which could in principle include a rich set of dependent measures (e.g. donations, event RSVPs, public comments, and so forth). Political actors use these very same tools in distributing their messages on the platform, and with the approach described above, researchers can employ fully randomized designs to study them.

There is no need to limit dependent measures to online behavior. By employing the randomization process described above, researchers can pair online trials with tried and true research methods in an exciting synthesis. As Broockman and Green (2012) describe, the age, gender, and location measures that form the basis for online-only randomizations appear in most American voter files right alongside individuals' telephone numbers. By pairing the clustered randomization algorithm described above with publicly available information on individuals in voter files, researchers can match participants in telephone polls to their treatment condition on the social network. In one study, for example, Broockman and Green thus partner with a candidate for state legislature and find Facebook ads designed to boost the candidate's name recognition appear to have minimal impact: follow-up telephone surveys found subjects in a treatment group that was targeted for an inundating advertising campaign were no more likely to know the candidate's name than subjects in a control group who were exposed to no advertisements. However, results from a second experiment also suggest that treated voters do recall seeing the ads at higher rates. As online advertising continues to mature, researchers thus may be able to employ this online cluster randomized design to experimentally investigate the impacts of mass communication with much greater ease than can be done with television, newspaper, and radio advertisements.

Cost

Not only can Facebook advertisements be randomized with ease, they are also surprisingly cheap:

ads often cost less than ten cents per 1,000 "impressions" (unique displays of the ad), or 0.0001 per exposure. This means that reasonably powered experiments can be conducted with a quite modest research budget: the three experiments reported by Ryan (2012), for instance, cost less than $1,000$ in total.

We leave it to future researchers to determine whether such a strikingly low price reflects a bargain for advertisers or a true reflection of the product's value.[4] (It may be diagnostic that advertisers are willing to pay thousands of dollars to expose individuals to television ads, but only pennies on Facebook ads with the same reach.) The most likely possibility, we suspect, is simply that Facebook ads are good at some things and not others. Perhaps, for example, they can engender spontaneous information search but are not likely to cause the kind of elaboration necessary to leave a lasting imprint on memory. Or, alternatively, perhaps the advertisements only pack a punch when they contain "social stimuli", such as in Bond and colleagues' most effective treatment. More work is clearly necessary to understand the impacts of this new presence in American political life, and with the tools described above researchers have many opportunities to explore its contours.

Conclusion: A Promising New Tool for Researchers

It is an exciting time to study political communication. Just as the ways that citizens engage in politics have exploded in number, so too have the approaches available to scholars who want to understand them. Along with other new resources such as Amazon.com's Mechanical Turk service (e.g., Berinsky, Huber, and Lenz 2012), we see social networking experiments as part of a rapidly expanding toolkit the promises to enrich our understanding of political participation and communication in myriad ways. Moreover, with so many Americans engaging in political deliberation online and being exposed to political messages, the medium itself should capture scholarly interest. We look forward to seeing how creative researchers will put these tools to use.[5]

<div align="center">

References

</div>

Arceneaux, Kevin. 2005. "Using Cluster Randomized Field Experiments to Study Voting Behavior." Annals of the American Academy of Political and Social Science 601: 169-179.

Berinsky, Adam, Gregory Huber, and Gabriel Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." Political Analysis 20(3): 351-368.

Blumberg, Stephen J. and Julian V. Luke. 2010. "Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2010." Available online at http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201106.pdf (accessed October 22, 2012).

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." Nature 489: 295-298.

---

[4]Readers interested in learning more about online advertising's impacts and pricing are directed to field experimental work by Lewis and Reiley (2012) and Lewis and Rao (2012).

[5]Both authors invite queries about the technical specifics of implementing randomized Facebook advertisements.

Broockman, David E. and Donald P. Green. 2012. "Can Facebook Advertisements Increase Political Candidates' Name Recognition and Favorability? Evidence from Randomized Field Experiments." Working Paper, University of California, Berkeley.

Chozick, Amy. 2011. "Television's Senior Moment," The Wall Street Journal (March 9, 2011). Available online at `http://online.wsj.com/article/SB10001424052748703559604576174983272665032.html` (accessed October 23, 2012).

Congressional Management Foundation. 2011a. "112th Congress Gold Mouse Awards." Available online at `http://congressfoundation.org/storage/documents/CMF_Pubs/cmf-112-gold-mouse-awards.pdf` (accessed October 22, 2012).

Congressional Management Foundation. 2011b. "SocialCongress: Perceptions and Use of Social Media on Capitol Hill." Available at `http://www.congressfoundation.org/storage/documents/CMF_Pubs/cmf-social-congress.pdf` (accessed October 28, 2012).

Facebook. 2012a. "One Billion - Key Metrics." Available online at `http://newsroom.fb.com/imagelibrary/downloadmedia.ashx?MediaDetailsID=4227&SizeId=-1` (accessed October 22, 2012).

Facebook. 2012b. "The Power of Facebook Advertising." Available online at `https://www.facebook.com/business/power-of-advertising` (accessed October 23, 2012).

Grimmer, Justin, Solomon Messing and Sean J. Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." Forthcoming, American Political Science Review.

Hampton, Keith N, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. "Social networking Sites and our Lives." Pew Research Center Internet and American Life Project. Available online at `http://pewinternet.org/Reports/2011/Technology-and-social-networks.aspx` (accessed October 22, 2012).

Hovland, C., A. Lumsdaine, and F. Sheffield. 1949. Experiments in Mass Communication: Studies of Social Psychology in World War II. Ann Arbor, Association for Consumer Research.

Lewis, Randall, and Justin M. Rao. 2012. "On the Near Impossibility of Measuring Advertising Effectiveness." Working Paper, Google.

Lewis, Randall, and David Reiley. 2012. "Does retail advertising work: Measuring the effects of advertising on sales via a controlled experiment on Yahoo." Working Paper, Google.

Panagopoulos, Costas, and Donald P. Green. 2008. "Field Experiments Testing the Impact of Radio Advertisements on Electoral Competition." American Journal of Political Science 52(1): 156-168.

Raine, Lee, Aaron Smith, Kay Lehman-Schlozman, Henry Brady, and Sidney Verba. 2012. "Social Media and Political Engagement." Pew Research Center Internet and American Life Project. Available online at `http://pewinternet.org/Reports/2012/Political-engagement/`

Summary-of-Findings/Overview.aspx (accessed October 22, 2012).

Ryan, Timothy J. 2012. "What Makes Us Click? Demonstrating Incentives for Angry Discourse with Digital-Age Field Experiments." Journal of Politics 74(4): 1138-1152.

Samuels, David J. and Cesar Zucco Jr. 2012. "Using Facebook as a Subject Recruitment Tool for Survey-Experimental Research." Working Paper, SSRN.

Sears, David O, and Jonathan Freedman. 1967. "Selective Exposure to Information: A Critical Review." The Public Opinion Quarterly 31(2): 194-213.

Simon, Scott. 2012. "Social Media A New Layer for Campaign Advertising" NPR.org: Weekend Edition Saturday. Available online at http://www.npr.org/2012/10/20/163296365/social-media-a-new-layer-for-campaign-messaging?sc=17&f=7 (accessed October 23, 2012).

Defend Elizabeth Warren
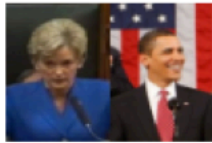my.elizabethwarren.com

Don't let Brown get away with his attacks. Stand with Elizabeth before it's too late!

Stop Karl Rove
action.kirstengillibrand.com

Special interest groups are spending millions to defeat Democrats in 2012. Fight back!

Not Again...
youtube.com

Michigan has seen this before, we do not need four more years of failed leadership.

Binders Full of Women?
emilyslist.org

Mitt Romney is out of touch with women & families. Keep his agenda out of the White House!
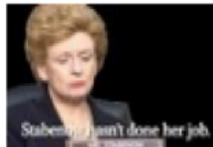
Gov Snyder: China Problem

Businessman Snyder outsourced American jobs to China. As a governor, he is doing it again!

GOP Holding Up Jobs
action.democraticgovernors.org

Pres. Obama has a plan for creating jobs, but Republicans are blocking it. Fight back now!

Third Term?
youtube.com

Does Debbie deserve a third term?

Voters Deserve to Know
youtube.com

Why won't Debbie Stabenow debate Pete Hoekstra?

Stop Scott Brown Now
my.elizabethwarren.com

Brown's not with us when it matters most. Help Elizabeth Warren hold him accountable!

Stand Against Mitt Romney
dscc.org

Don't let Mitt get away with his dishonest debate attacks. Sign up here and fight back!

Figure 1. Some political ads observed on Facebook by Ryan in October of 2012. The ads seem to reflect broad targeting strategies on the part of candidates, since they come from both Republican and Democrat sources, as well as from both inside and outside of Michigan (Ryan's state of residence).

	Male	Female
18-25	31.8%	26.9%
26-39	11.8	10.4
30-39	21.8	21.2
40-49	15.8	17.2
50-59	10.2	13.5
60+	8.7	10.8
	100.0	100.0
N	69.9M	83.3M

Table 1. Source: Authors' calculations using Facebook's advertising tool, which reports the number of people by hypothetical demographic targets. Facebook permits users to join once they are 13, but users between the ages of 13 and 17 are excluded from analysis.

Age	Geography	Gender	Condition
18	Ann Arbor	Male	B
18	Ann Arbor	Female	B
18	San Francisco	Male	A
18	San Francisco	Female	B
19	Ann Arbor	Male	A
19	Ann Arbor	Female	A
19	San Francisco	Male	A
19	San Francisco	Female	B
⋮	⋮	⋮	⋮
64	San Francisco	Male	B
64	San Francisco	Female	A

Table 2. This experiment would have 47 (age) × 2 (Ann Arbor/ San Francisco) × 2 (Gender) = 188 clusters.

Using Dynamic Process Tracing to Study Decision Making: A Laboratory Research Tool for the Scholarly Community

David P. Redlawsk

Department of Political Science and Eagleton Institute of Politics, Rutgers University

redlawsk@rutgers.edu

Over the years, Rick Lau and I have argued that the dynamics of decision making cannot be studied effectively with static approaches, such as cross sectional survey research. Instead, what is needed is a methodology designed from the start to follow decision makers as they make their evaluations and choices. It's all well and good to ask people to recall what they looked at, or how they felt about the information they acquired prior to a decision, but human beings are notoriously unable to recall many of the specifics they encounter on the way to a decision. As is now well-established, evaluations of social objects - like candidates - often proceed online, with the details that informed the evaluation discarded or at least relatively inaccessible after the fact (Lodge, McGraw, and Stroh, 1989, Kunda, 1987, Redlawsk, 2001.) This is not to say that memory plays no role (Lau and Redlawsk, 2006; Civettini and Redlawsk, 2009), but even to the extent that it does, memory recall is rarely, if ever, a perfect representation of the information that informed the decision. Whether evaluation and choice are online processes, memory processes, or some hybrid of the two, the best way to study the dynamics of a decision is as it happens. Obviously that position suggests that laboratory studies are an important tool in decision making research.

This perspective on decision making generally, and voter decision making specifically, led Lau and I to develop the Dynamic Process Tracing Environment (DPTE) used by us in a wide range of experimental studies on voter decision making. The initial DPTE research, summarized in our 2006 book, How Voter Decide (Cambridge University Press) and a number of papers on cognitive heuristics, online processing, and motivated reasoning, opened new perspectives on limited information processing versus rational decision making models, the role of affect in driving motivated reasoning, and of course, the concept of "Voting Correctly" (Lau and Redlawsk, 1997). This short report briefly describes some of my own and colleagues' research that has been stimulated by the availability of this tool, and describes how researchers can get access to this tool which is now available online.

Much of my decision making research beyond the initial information processing studies has focused on emotional responses to candidates, using DPTE experiments to understand the interplay between emotional and cognitive processing as voters encounter a campaign. Early on I was intrigued by the findings of Milt Lodge and his colleagues on online processing (Lodge, McGraw, and Stroh, 1989; Lodge, Steenbergen, and Brau, 1995) and motivated skepticism (Taber and Lodge, 2006; also see a new issue of Critical Review 24(2): 2012, devoted to their theory) but also unsure of how they might apply to the dynamics of a political campaign, which are always changing and involve multiple players. At the same time, George Marcus and his colleagues (Marcus, Neuman, and MacKuen, 2000) developed a theory of affective intelligence, which argued that emotions - anxiety in particular - could lead to better voter decision making. Together these theories suggested that affect could both improve voter

decision making (AI) or lead it astray (motivated skepticism).

Both theories provide important insights into decision making and pushed the field to go beyond the purely information processing paradigm that had developed in the 1980's and 1990s. But both were developed using static techniques, either survey research via ANES data for Marcus or relatively simple laboratory studies of a single candidate outside of an election context (Lodge and colleagues' online processing studies.) Having a new tool at hand - the earliest version of Dynamic Process Tracing - I sought to understand if there were contexts in which the two approaches could be reconciled, if not theoretically, at least empirically. In particular, it seemed necessary to test online versus memory processing, and the idea that rising anxiety would lead to increased information search and potentially better decisions, in the complex context of a campaign with competing candidates.

In brief, DPTE has been used to confirm that online processing is a default processing mode for voters in a campaign environment but that memory plays an important role in the process of making a choice, as opposed to simply developing an evaluation (Redlawsk, 2001, Redlawsk, 2006; Lau and Redlawsk, 2006.) Given the presence of online processing, Lodge and Taber (2000) argued that voters might be motivated reasoners, something a DPTE study confirmed (Redlawsk, 2002.) But at the same time, evidence appeared in our process tracing data for a process that looked like affective intelligence, under conditions of increasing threat to pre-conceived preferences (Redlawsk, Civettini, and Lau, 2007). Are these contradictory? That is, can voters simultaneously become more careful processors, and more knowledgeable about their choices, and yet be motivated to maintain an existing evaluation even when that evaluation is no longer supported by the facts? One possibility is that there is in fact what we call an "affective tipping point", a point at which even the strongest motivated reasoner recognizes reality and rather than continuing to strengthen a positive evaluation in the face of negative information, finally begins to update more accurately for the remainder of the campaign (Redlawsk, Civettini, and Emmerson, 2010). Motivated reasoning may initially be triggered when unexpected (negative) information about a liked candidate is encountered. But if more negative information accumulates, maintaining a positive evaluation appears to be anxiety-provoking, leading to the potential for reassessment and a readjustment.

What all of these studies have in common is their methodology, unabashedly laboratory studies, but designed in a complex environment that can mimic the potential chaos of a campaign environment. By conceiving of evaluation and choice as a process, rather than a static thing, we can better account for how candidate evaluations are developed and updated and the joint role of emotion and cognition.

During the past few years my primary focus has been on the development of DPTE as a flexible research tool for laboratory and online studies of decision making processes. The National Science Foundation has funded this software development effort, which has resulted in a system now available to the research community as a whole. The basic paradigm is that the system can be programmed to present information in video, image, or text formats, which subjects may choose to examine as they move towards making evaluations and choosing between alternatives. The experimenter has a wide range of options to create a flexible process-tracing experiment, where every action taken by the subject is recorded, enhancing our ability to understand decision-making as it happens. Moreover, recently

added features allow the development of social experiments, where participants can choose to share, like, dislike, or comment on the information they see during an experiment.

While Rick Lau, our students and early adopter colleagues, and I have mostly applied it to voting studies, DPTE can be used to study any type of decision-making process in which information flows over time. Examples include studies of decision making in times of crisis, studies of the effects of information sources and credibility, studies of public policy debates and referenda, and many others. This tool provides an experimental environment in which decision makers encounter information that comes and goes over time, make active choices of what information to examine and what to ignore, are possibly exposed information they did not select, and ultimately choose between alternatives, or perhaps make no choice at all. Our hope is that this new methodology will enhance the ability of those doing behavioral decision making research to test hypotheses about how people process information - cognitively and emotionally - over time.

Any researcher wishing to use the software may acquire a user ID by navigating to the site: http://www.processtracing.org and clicking on the "Request Researcher ID" link. You will be asked to provide some information and upon approval a link will be sent to allow you to set up your account. The user guide for the software is also available at this location, by clicking on the "Help" link. We have recently begun a blog to describe how the system is being used and to provide tips and techniques for designing experiments, which can be accessed at http://dynamicprocesstracing.wordpress.com.

Ultimately, our goal is to make a new research tool available to the scholarly community that allows behavioral researchers to conduct sophisticated computer-based experiments on human decision making and social behavior in dynamic settings without devoting years to acquiring the requisite programming skills. It is our belief that while we began this project focused on how voters use information to evaluate candidates and make a voting decision; a much wider array of studies can be carried out using the DPTE methodology. Examples include studies of decision making in times of crisis, studies of the effects of information sources and credibility, studies of public policy debates and referenda, and many others. The key point of this methodology is to provide an experimental environment in which decision makers encounter information that comes and goes over time, make active choices of what information to examine and what to ignore, and ultimately choose between alternatives, or perhaps make no choice at all. Our hope is that this new methodology will enhance the ability of those doing behavioral decision making research to test hypotheses about how people process information - cognitively and emotionally - over time.

References

Civettini, Andrew and David P. Redlawsk. 2009. Voters, Emotions, and Memory. Political Psychology 30(1): 125-151.

Kunda, Z. 1990. The Case for Motivated Reasoning. Psychological Bulletin, 1083: 480 - 498.

Lau, Richard R. and David P. Redlawsk. 1997. Voting Correctly. American Political Science Review, 91 (September): 585-598.

Lau Richard R. and David P. Redlawsk. 2001. Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making. American Journal of Political Science 45(October): 951-971.

Lau, Richard R. and David P. Redlawsk. 2006. How Voters Decide: Information Processing during an Election Campaign. New York: Cambridge University Press.

Lodge, Milton, McGraw, Kathleen., and Stroh, Patrick 1989. An impression driven model of candidate evaluation. American Political Science Review 83:399-419.

Lodge, Milton, Marco R. Steenbergen, and Shawn Brau. 1995. The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation. American Political Science Review, 89(June): 309-326.

Marcus, George, W. Russell Neuman, and Michael McKuen. 2000. Affective Intelligence and Political Judgment. Chicago: University of Chicago Press.

Redlawsk, David P. 2001. You Must Remember This: A Test of the On-line Model of Voting. Journal of Politics, 63(February): 29-58.

Redlawsk, David P. 2002. Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making. Journal of Politics 64(November): 1021-1044.

Redlawsk, David P. 2006. Motivated Reasoning, Affect, and the Role of Memory in Voter Decision-Making. In David Redlawsk (Ed.) Feeling Politics: Emotion in Political Information Processing. New York: Palgrave Macmillan.

Redlawsk, David P., Andrew Civettini, and Karen Emmerson. 2010. The Affective Tipping Point: Do Motivated Reasoners ever "Get It"? Political Psychology 31(4): 563-593.

Redlawsk, David, Andrew Civettini, and Richard R. Lau. 2007. Affective Intelligence and Voter Information. In W. Russell Neuman, George Marcus, Michael MacKuen and Ann Crigler (Eds.) The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior. Chicago: University of Chicago Press.

Taber, C. S. and Lodge, M. 2006. Motivated Skepticism in the Evaluation of Political Beliefs. American Journal of Political Science 50: 755-769.

Measuring Sensitive Attitudes in Developing Countries: Lessons from Implementing the List Experiment

Eric Kramon and Keith R. Weghorst
kramon@ucla.edu and keith.weghorst@ufl.edu
UCLA/Stanford University and University of Florida

Lessons on using the List Experiment in Developing Countries

   With the spread of elections to new democracies and undemocratic regimes has risen attention towards electoral violence, fraud, and clientelism. The corresponding growth in related scholarship is striking, as shown in figure 1. Conventional tools for collecting citizens' viewpoints about these phenomena, however, are woefully inadequate. In sub-saharan Africa, where up to 45% of respondents

suspect public opinion surveys are carried out by the government,[6] incentives to lie about selling votes for cash or engaging in political violence are often too great to gather reliable public opinion data.

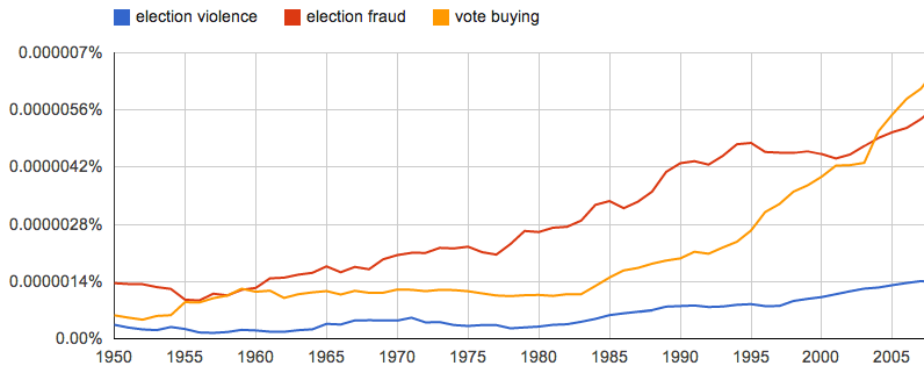Figure 1: **Increasing Popularity of Hard to Measure Topics**

Figure 1 shows the percentage of books in the English-language corpus from 1950 to 2008 that contain the bigrams "election violence," "election fraud," and and "vote buying." Created with Google's nGram Viewer.

The "list experiment" is an increasingly popular alternative to direct question formats that offers respondents more anonymity, but at the cost of additional complexity. What remains to be seen is whether assumptions about the steps individuals take to answer such questions are valid in low-income environments, where both respondent literacy and numeracy rates are low and sensitivity biases about these topics abound. In what follows, we discuss our attempts to validate these assumptions in developing countries. We offer evidence of a central problem with list experiments in these settings: respondents *satisfice* by skipping key steps in the response process, steps that are key to collecting valid list experimental data. We then discuss two modifications of the list experiment that address the problem of satisficing and offer a general set of lessons for carrying out list experiments in developing country settings.

What are list experiments?

The list experiment (alternately called the "item count technique") combines elements of randomization and aggregation to draw inference about sensitive attitudes and behaviors in a way that cannot later be directly linked to respondents. In its simplest form, the list experiment randomly assigns respondents into treatment and control groups. Control group respondents are given a list of statements about non-sensitive attitudes or activities that are not subject to response bias (typically three or four items). Treatment group respondents receive a list with the same non-sensitive items and the sensitive item of interest. All respondents are instructed to respond with the total number of applicable items. Random assignment implies that the number of applicable non-sensitive items will be the same for treatment and control groups. Provided that the addition of the sensitive item does not

---

[6]Based on Afrobarometer's 2008 survey in 20 countries in the sub-continent, with an additional 15% suspecting political institutions like the parliament and political parties.

impact how respondents consider the non-sensitive ones ("no design effects") and respondents are truthful ("no liars") the difference of means between treatment and control groups provides an unbiased estimate of the proportion of the population to whom the sensitive item applies (Blair and Imai, 2012). The technique is evidenced to successfully reduce downward biases about sensitive political behaviors and attitudes, from voter turnout (Holbrook and Krosnick, 2010), the influence of racism and sexism on candidate preferences (Kane et al., 2004; Kuklinski et al., 1997; Martinez and Craig, 2010; Rayburn et al., 2003; Streb et al., 2008) views on political violence (Weghorst, 2012) and experience with vote-buying (Gonzalez-Ocantos et al., 2011). In spite of these encouraging developments in the use of list experiments, there remain many unresolved questions about their ability to provide unbiased estimates of sensitive behaviors.

Learning by experience with list experiments in sub-Saharan Africa

We have carried out a total of eleven list experiments in Kenya and Tanzania from 2009-2012. They have been in different regions of the countries with distinctive population subgroups (citizens and legislators) and in diverse environments, including face-to-face interviews in household and public settings and self-administered instruments in classroom and "take-home" settings.[7] To date, we have found that the list experiment can effectively reduce sensitivity biases about vote-buying and views on electoral fraud, political violence, and government repression. Our research has also allowed us to explore many empirical properties of the list experiment in developing country settings, where the popularity of the technique has increased greatly over recent years.

An important lesson from these projects has been that the list experiment obliges respondents to expend more cognitive effort to accurately provide answers than do conventional question formats. We have further observed that the additional necessary effort is much greater in environments with low levels of education, literacy, and numeracy. Like all survey questions, each list experiment item requires careful consideration of its meaning, recalling all information necessary to evaluate it, and synthesizing the information into a summary judgment used to provide a response (Tourangeau, 1984). Unlike direct questions, list experiments also introduce unfamiliar instructions and ask respondents to hide individual item evaluations and instead internally aggregate judgments across many statements into a numerical value. If respondents view these tasks together as too burdensome or taxing, the list experiment can induce response effects. Our research has culminated into an exploration of a specific response effect known as "satisficing." Satisficing occurs when an interviewee skips one or more step required to fully answer a survey question and only puts forth enough effort to provide a response he or she deems acceptable or satisfactory (Krosnick, 1991). For list experiments, satisficing could occur if, instead of extensively evaluating each item and counting in their heads, respondents simply report a numerical value they think is reasonable.

Satisficing should be observable when responses to direct questions about list items do not match the information provided about those items in a list experiment format. Detecting satisficing with list experimental data, however, is difficult due to the sensitivity biases of direct questions and potential complexity of control items. In our most recent round of data collection, we developed a test of

---

[7]More information on these list experiments can be found at www.ericjkramon.com and www.keithweghorst.com

satsificing using simple items that facilitated easy recall and would not be subject to sensitivity biases. We generated two lists of entirely non-sensitive activities which respondents readily remember. The first asked how many of five foods were consumed in the week prior and the second how many of five numeracy related activities were performed in the last month. The latter included things like sending money over mobile phone (mPESA) and selling goods for profit. Enumerators administered these items in the conventional list experiment format and then, about five minutes later in the survey, asked respondents about each of the items as direct "Yes/No" questions. The results suggested that list experiments in this setting—urban slums in Kenya—are rife with satisficing. For both lists, more than 40% of respondents' list experiment responses did not match the true value identified through direct questions. Figure 2 illustrates the prevalence of satisficing, measured as the difference in the number of applicable items elicited through the two formats. Zero represents a match. Negative values show under-predictions of the list experiment and positive values over-predictions.

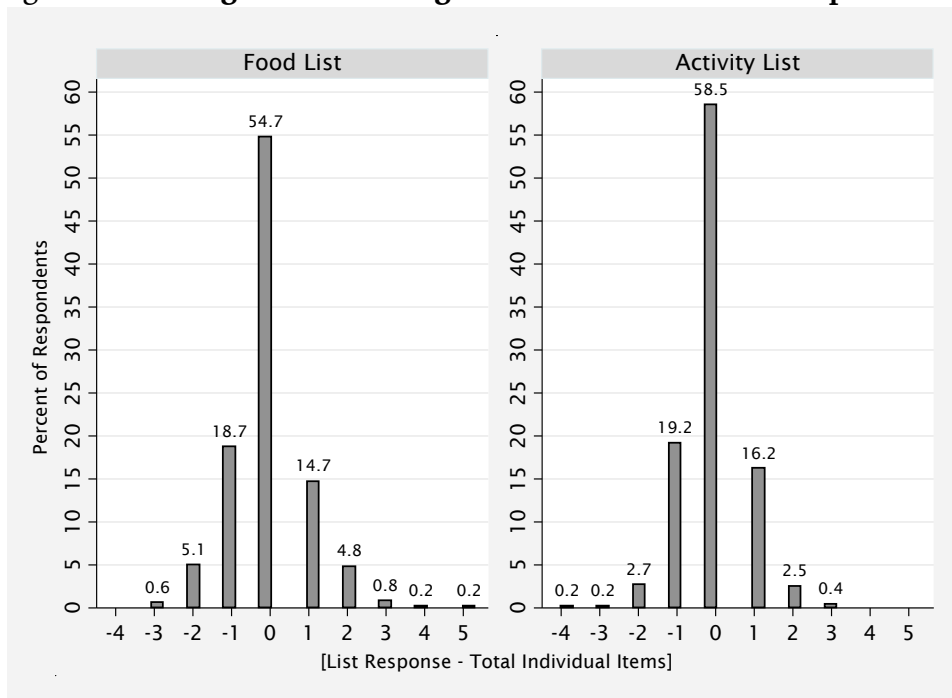Figure 2: **Testing for Satisficing with non-sensitive list experiments**

Figure 2 shows the frequency distribution of the difference between the number of foods and activities indicated by respondents for the list experiment format and direct questions. Instances of of non-matches were slightly skewed towards under-reporting, with 24% and 22% of interviewees reporting a lower number of items for the list experiment than for the direct question, respectively.

Analysis of what determines non-matches revealed that each additional level of educational attainment is associated with a 20% higher probability of matches between the two formats for both

the food and activity lists. We want to be very clear that we do not think these list experiments failed for lower education subgroups because of respondent intelligence, particularly because educational attainment in developing countries is frequently determined by societal factors—gender, birth order, family income, etc. Rather, we view education as a crude indicator of experience with higher order cognitive tasks and the cognitive sophistication required to optimally answer list experiment questions. When respondents lack experience using these skills, the effort a respondent must put forth appears simply too great to obtain accurate answers. The same holds for numerical literacy: using the total number of numeracy related activities as an indicators of this, we find respondents who engage in fewer numerical activities are less likely to match between list experiments and direct questions for the food.

To address the satisficing problem, we modified list experiments procedures in two ways. Our modifications aimed to (1) make each item's meaning quickly clear to respondents, (2) ease aggregation by allowing private visual cues and note-taking, (3) build on familiar tools used to communicate about politics, and (4) encourage full effort by making list experiments more active and engaging for respondents. The first modification provided respondents with a laminated copy of the list and a dry erase marker. Enumerators showed respondents how to use a dry erase marker on the laminate—ticking the sheet and erasing marks with a fingertip—handed the materials over, turned 90 degrees away, and read the lists. Our second innovation also provided respondent handouts and had the enumerator turn 90 degrees away from the respondent. The second set of handouts contained cartoons we commissioned that corresponded with each item on the list. Cartoons are commonly used by NGOs and development agencies—including for public campaigns to reduce political violence and fraud—in settings of low literacy and educational attainment, where we expect satisficing to be most prevalent. They are also captivating and more familiar to respondents than the list experiment. Figure 3 is one such cartoon we used in Kenya, corresponding with the statement "If another tribe tries to steal an election, it is justified to use violence to try to stop them" and shows a thief attempting to manipulate presidential election results, while others look on.[8]

We presented respondents with the list experiment in its standard form and at a different part of the survey as one of the modifications to test their efficacy.[9] For the full sample, respondents reported that violence was justified at lower rates with the conventional list experiment format than they did the direct question, the opposite of what we would expect given the downward sensitivity biases present with direct questions. By comparison, the tick and cartoon modifications yielded higher prevalence rates than the conventional list experiment and the direct question. We interpret this as the modifications enhancing the performance of the list experiment. The improvements for the cartoon modification were most pronounced for low education and low numeracy respondents, while those for the tick procedure occurred more broadly across education and numeracy subgroups.

We also think we have addressed the issue of satisficing, which is supported by figure 4. This figure presents results for the "satisficers," whose list experiment and direct question responses did not match

---

[8]In order to ensure using cartoons did not somehow trivialize violence—an ethically troubling issue that also could reduce sensitivity biases—we used the word "Mwizi," a very serious Kiswahili term for "thief" which is rarely, if ever, used with levity.

[9]We randomized whether a respondent received the standard or modified form first and distracted respondents with additional and unrelated questions in-between, including the food and activity lists.

Figure 3: **Cartoon Corresponding with Violence Statement**
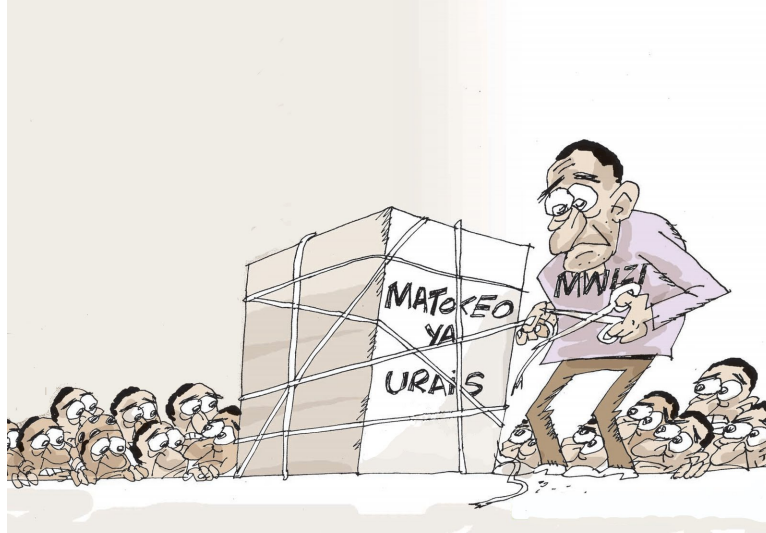
for both the food and activity lists. It provides evidence of satisficing (that the standard list experiment under-predicts the sensitive behavior) but that the cartoon and tick modifications show higher prevalence of the sensitive attitude.

Lessons for Carrying Out List Experiments

We now offer suggestions for designing and implementing list experiments. We do not discuss growth in analytical tools for list experimental data, but laud contributions that include double-list experiment designs that reduce inefficiency of group-level comparisons (Droitcur et al., 1990; Glynn, 2010), ways to detect violations of certain list experiment assumptions (Blair and Imai, 2012) and statistical packages that allow multivariate (Corstange, 2009) and individual-level analyses (Blair and Imai, 2012).

1. Low-cost, seasoned tools over high-tech, unfamiliar solutions

Our modifications were very low cost—dry erase markers, printing and laminating handouts, and a few hundred dollars in cartoon commissions. Instructing enumerators to use them was easy and they resonated with respondents. The cartoon modification built on NGOs and development agencies success in communicating complex and unfamiliar topics. We believe an alternative like electronic devices that allow respondents to privately enter responses would have dramatically increased project costs without actually addressing the underlying factors contributing to list experiment breakdown. For a technique already difficult enough to induce satisficing, new technologies can exacerbate the unfamiliarity of list experiments.

2. Make privacy more credible

In face-to-face interviews, offering respondents a way to privately tabulate items may reduce satisficing. However, potential fears that a pen-stroke or other more subtle physical responses are detectable when listening to items must be addressed by making privacy real to respondents. In our

Figure 4: **Estimates of Agreement with the Sensitive Item Among "Satisficers"**
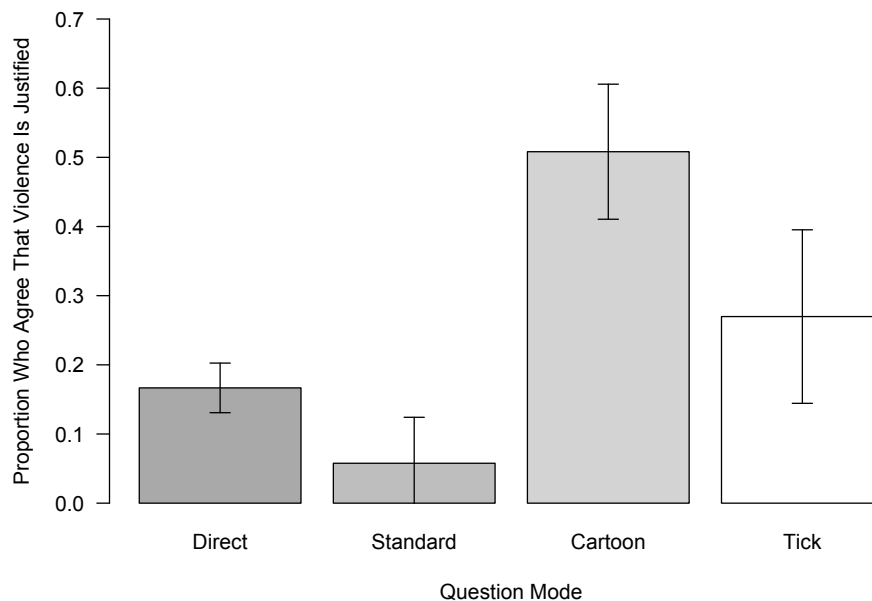

Figure 4 shows the estimated agreement with the statement "If another tribe tries to steal an election, it is justified to use violence to try to stop them" for the 108 individuals who did failed to match on both the food or activity lists presented in figure 2. It compares the direct question, standard list experiment design, and our cartoon and tick modifications.

experience, the combination of visual aids with and enumerators turning 90 degrees away worked. Creating a secure environment requires understanding privacy concerns of respondents have and carries a strong ethical obligation to ensure the technique hides an individuals view on the sensitive item.

3. Find instructions that work

Respondent instructions for introducing the list experiment are extremely important, as they must effectively discourage respondents from identifying which specific items apply to them. Firstly, work carefully with enumerators to develop simple instructions that interviewees easily internalize. Secondly, pre-test instructions with respondents from the target population. You may find first administering an artificial list is necessary for respondents to understand how the list experiment works. Thirdly, in addition to a survey's standard definitions and feedback manuals, develop enumerator protocols for when respondents need additional instructions to understand the list experiment. In our experience, enumerators who try to explain the list experiment to respondents without guided instructions frequently point to the secrecy the technique provides (e.g., "A special kind of question that hides your individual answers about vote-buying but lets us see much there is in Kenya in general"). We caution

against enumerator protocols that specifically note how secrecy should lead to higher reporting rates of the sensitive attitude, as it could induce Hawthorne effects where treatment group respondents report higher numbers of items as an artifact of how they think the technique "should work."

4. Keep control list items short and punchy

The effort list experiments require to optimally answer can lead to satisficing. Keeping control items brief and easily interpretable can reduce this effort substantially. Using items (including the sensitive one) in similar topical and temporal areas of respondents' memories can make recall less laborious. Successful list experiments we have carried-out employed non-sensitive items that were only a few words and were accessible enough to elicit quick, "knee-jerk" reactions from respondents.

5. Field test non-sensitive items

Learn a lot about people's views towards non-sensitive items before using them on list experiments. It may be possible to start with pre-existing data—we consulted Afrobarometer surveys to anticipate response patterns about prospective non-sensitive items. Selecting items with predictable response patterns can reduce variance of control lists and, by extension, the non-sensitive items in the treatment group. Well documented field tests of non-sensitive items are also important to determine what items "work" and "don't work" for researchers who want to carry out list experiments about similar topics or in similar settings. To our knowledge, the most extensive annotation of list experiment designs, including control and treatment list items can be found at Kosuke Imai's website.

6. Standards for list experiment design are growing

- *6.1 Avoid Ceiling and Floor Effects*

  A control list where zero or all items could plausibly apply to a respondent cannot ensure the anonymity the design attempts to provide respondents. Known as "floor" and "ceiling" effects (Glynn, 2010), consequent advice is to include on control lists at least one item that very few respondents will find applicable and one that nearly all find applicable.

- *6.2 Whenever possible, ask the direct question*

  Asking a direct question allows us to assess how well the list experiment reduces sensitivity biases. Because list experiments are less computationally efficient than direct questions, doing so better elucidates trade-offs of using one over the other. Direct questions also help us identify "smoking gun" failures, when the list experiment under-estimates the prevalence of a sensitive attitude compared to a direct question. The most comparable direct questions will use a single statement with "Yes/No" response options, with a question prompt matching the list experiment instructions (comparing "Is the following true for you? I agree that [sensitive statement]" with "How many of the statements do you agree with...?") Ethical considerations may prohibit asking the direction question, but we echo Blair and Imai (2012) in reiterating the importance of doing so when at all possible.

- *6.3 Consider other techniques to triangulate list experiments*

  There are many experimental and non-experimental techniques designed to reduce sensitivity biases present with self-reported behaviors and attitudes. Each approach has certain trade-offs.

Comparing list experiment results to other alternatives can inform for what topics and in which settings it is most effective. In a study we carried out in Tanzania, we learned that the list experiment was far preferable to the more complex randomized response technique (see: Warner, 1965) and that the latter approach performed poorly even amongst university educated populations there (Weghorst, 2012). Others have compared list experiments to endorsement experiments and found similar results through the two techniques, lending greater confidence to the list experiment findings (Blair, Imai, and Lyall 2012).

7. Many puzzles still remain

- *7.1: Are there item order effects?*

  Item order and placement may matter for several reasons. Firstly, if respondents satisfice, they may pay closer attention to certain items simply based on their position in a list. A sensitive item placed at the beginning of a list might change how respondents consider subsequent items. Randomizing order can help detect ordering effects, but we recommend against randomizing both control and treatment lists. We suggest randomizing control list order and randomizing the position of the treatment item amongst the fixed control list order to keep control and treatment lists as similar as possible. Potential list order and sensitive item placement effects offer a promising avenue for future research.

- *7.3: Open or Closed Control Lists?*

  Asking respondents to elaborate which control items are applicable can improve inference (Corstange, 2009) but not if interviewees respond differently when elaborating versus aggregating themselves (Flavin and Keane, 2009; Tsuchiya and Hirai, 2010). Different aggregation instructions for control group could mean the treatment group is simultaneously treated with the sensitive item and an alternative aggregation procedure, but this needs to be addressed more extensively in future work.

- *7.2: How many non-sensitive items?*

  Control list length may impact the aggregated number respondents of items report, regardless of the applicability of non-sensitive items (Nickerson and de Jonge, 2012; Tsuchiya et al., 2007). Reasons why may lie in satisficing. Studying possible list-length effects is important for designing control lists of ideal length and for list experiment inference because control and treatment lists are different lengths.

Conclusion

List experiments present a promising way to elicit sensitive attitudes without suffering from biases of self-censorships. As their appeal for studying political phenomena in the developing world, we encourage scholars to further evaluate their performance in these settings and to continue developing design and analysis tools that improve inferences we make from list experiment data.

References

Blair, Graeme and Kosuke Imai. 2012. Staistical Analysis of List Experiments. Political Analysis 20(1): 47- 77.

Blair, Graeme, Kosuke Imai and Jason Lyall. 2012. Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan. Under Review.

Corstange, D. 2009. Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. Political Analysis 17(1): 45-63. Droitcour, J., Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher and Tena M. Ezzati. 1991. The item count technique as a method of indirect questioning: A review of its development and a case study application. Measurement Errors in Surveys pp. 185-210. Flavin, P. and M. Keane. 2009. How angry am I? Let me count the ways: Question format bias in list experiments. Unpublished Manuscript.

Glynn, Adam N. 2010. What Can We Learn With Statistical Truth Serum?: Design and Analysis of the List Experiment. Unpublished Manuscript.

Gonzalez-Ocantos, E., Chad Kiewet de Jonge, Carlos Melendez, Javier Osorio and David W. Nickerson. 2011. Vote buying and social desirability bias: Experimental evidence from Nicaragua. American Journal of Political Science 56(1): 202-17.

Holbrook, Allyson L. and Jon A. Krosnick. 2010. Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique. Public opinion quarterly 74(1): 37-67.

Kane, James G, Stephen C Craig and Kenneth D. Wald. 2004. Religion and Presidential Politics in Florida: A List Experiment. Social Science Quarterly 85(2): 281-93.

Krosnick, Jon A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. Journal of Cognitive Psychology 5: 213-36.

Kuklinski, James H., Michael .D. Cobb and Martin Gilens. 1997. Racial attitudes and the New South. Journal of Politics 59: 323-349.

Martinez, Michael D. and Stephen C Craig. 2010. Race and 2008 Presidential Politics in Florida: A List-Experiment. The Forum 8(2): 1-14.

Nickerson, David W. and Chad Kiewet de Jonge. 2012. Artificial Inflation or Deflation? Assessing the Item Count Technique in Comparative Surveys. Under Review.

Rayburn, Nadine Recker, Mitchell Earleywine and Gerald C. Davison. 2003. An Investigation of Base Rates of Anti-Gay Hate Crimes Using the Unmatched-Count Technique. Journal of Aggression, Maltreatment and Trauma 62(2): 137-53.

Streb, Matthew J., Barbara Burrell, Brian Fredrick and Michael A Genovese. 2008. Social Desirability Effects and Support for a Female Presidential Candidate. Public Opinion Quarterly 72(1): 76-89.

Tourangeau, Robert. 1984. Cognitive Sciences and Survey Methods. In T. Jabine, M. Straf, J. Tanur and Robert Tourangeau (Eds.) Cognitive aspects of survey methodology: building a bridge between disciplines. Washington, DC: National Academy Press pp. 73-199.

Tsuchiya, Takahiro and Yoko Hirai. 2010. Elaborate Item Count Questioning: Why Do People Underreport Item Count Responses? Survey Research Methods 4(3): 139-49.

Tsuchiya, Takahiro, Yoko Hirai and Shigeru Ono. 2007. A study of the properties of the item count technique. Public Opinion Quarterly 71(2): 253-272.

Warner, Stanley L. 1965. Randomize Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association 60(309): 63-9. Weghorst, Keith R. 2012. Uncovering Sensitive Political Attitudes with List Experiments and Randomized Response: A Survey Experiment Assessing Data Quality in Tanzania. Working paper.

Review of *The Social Citizen* By Betsy Sinclair

Anand Edward Sokhey
University of Colorado at Boulder
Anand.Sokhey@Colorado.EDU

Over the last 30 years, our understanding of how interpersonal networks affect politics in the mass public has been advanced time and again by Huckfeldt and colleagues (e.g., Huckfeldt and Sprague 1995); the question of how our friends, family and acquaintances shape our opinions and behaviors remains a topic of scholarly debate, as it has varied implications for democratic functioning (e.g., Mutz 2006).

Of course, research on social influence has also been pestered by problems, and chief among these has been the thorny issue of self-selection. When individuals choose others that mirror them on various characteristics (homophily), should we believe causal claims about influence? For decades, scholars have had to address this concern through less than optimal designs, resulting in skepticism and lingering questions.

Enter The Social Citizen. Building on recent efforts using temporal (e.g., Klofstad 2011) and experimental approaches (e.g., Nickerson 2008), Betsy Sinclair systematically addresses the threats to inference that have long plagued the literature. Bringing together field experiments, original survey data, and careful analytic techniques, Sinclair demonstrates that social factors sway turnout, participation (political contributions), candidate choice, and even party identification. After situating the broader effort in the existing literature (and providing a theoretical overview), chapter two presents the results from two original field experiments on turnout. These two "cuts" at social influence provide evidence for the "social component of voting" (32), one of which emerges via an innovative, multi-level field experiment designed to gauge spillover effects (Sinclair, McConnell and Green 2010). Individuals are more likely to turnout when a local canvasser delivers a mobilization message; they are more responsive to mobilization treatments when they reside with others who have internalized participatory norms (by voting consistently).

In chapter three, Sinclair shifts her attention to a different form of participation - donating to campaigns. Focusing on an Illinois congressional district, she uses publicly available data on political contributions to "back out" a (whole) network, connecting individuals by their common giving to campaigns or organizations. With this network in place, she then calculates centrality scores for individuals, and links these - which gauge public visibility - to dollar amounts. The idea presented is that giving is a social act that is fundamentally about the observance of and adherence to norms. This point is further underscored via an original survey of donors and a number of in-person interviews.

In the final two empirical chapters, Sinclair turns to original and established survey data, looking at candidate choice (chapter four) and that most fundamental of behavioral concepts, party identification (Campbell et al. 1960) (chapter five). Leveraging temporal information (in the form of panel data) and careful modeling strategies, she demonstrates that social networks - measured via ego-centric data - exert influence on these preferences. A variety of robustness checks are presented, and the constellation of results provides compelling evidence for causal inferences.

The Social Citizen makes a contribution by demonstrating - definitively - that network effects are real. On this point alone the book should (and will be) cited heavily. That said, there is another contribution lurking behind the presentation of clean designs and "teched-up" methodological tools. Across chapters and dependent variables, Sinclair conceptualizes networks and social influence in a variety of ways. However, the idea is not just to demonstrate that networks matter, but to demonstrate how they matter. All analyses are united by a focus on mechanisms, and specifically, on a sustained effort to adjudicate whether social influence operates via explicit information transmission, or social pressures (that is, the expression, acknowledgement, and adherence to behavioral norms). The accumulation of evidence points to pressure rather than information as the mechanism of influence. This is an important finding, and one that will spur additional scholarly inquiry.

While The Social Citizen offers many advances and insights, it will leave readers with a few questions. Some of the evidence provided is more suggestive than conclusive (e.g., a portion of the experimental results in chapters two; parts of chapter three). Likewise, one wonders about the robustness of the findings to alternative measurement strategies. For example, in the survey-based results, would different conceptualizations and operationalizations of disagreement (e.g., Klofstad, Sokhey and McClurg 2012) yield similar results? Would better measures of sophistication still yield null results for information-based theories of social influence? Finally, and most broadly, what are the downsides to a social citizenry? Sinclair begins to address this (and related "big picture" points), but the narrative is less-developed on this front.

Still, The Social Citizen emerges as a thorough and thoughtful effort. The book is simultaneously accessible and technical; it responds to an established literature at the same time that it leads it in new directions. It may be premature to say that Sinclair has put the "causal effects fight" to rest when it comes to the study of social influence. However, she has certainly landed some blows, and in doing so, has urged us to consider some critical questions.

References

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. Voting. Chicago: University of Chicago Press.

Campbell, Angus, Phil Converse, Warren Miller, and Donald Stokes. 1960. The American Voter. New York: Wiley.

Huckfeldt, Robert and John Sprague. 1995. Citizens, Politics, and Social Communication. New York: Cambridge University Press.

Klofstad, Casey A. 2011. Civic Talk: Peers, Politics and the Future of Democracy. Philadelphia, PA:Temple University Press.

Mutz, Diana. 2006. Hearing the Other Side: Deliberative Versus Participatory Democracy. Cambridge: Cambridge University Press.

Nickerson, David W. 2008. Is Voting Contagious? Evidence from Two Field Experiments. American Political Science Review. 102(1):4957.

Sinclair, Betsy, Margaret A. McConnell, and Donald P. Green. 2010. Detecting Social Networks: Design and Analysis of Multi-level Experiments. Paper presented at the Third Annual CESS NYU Experimental Political Science Conference, February 5-6, 2010.

Review of *A Model Discipline: Political Science and The Logic of Representations* By Kevin Clarke and David Primo

Jonathan Woon
University of Pittsburgh
woon@pitt.edu

Is political science really science? We hear this question from physicists, chemists, and even computer scientists. We also hear it from politicians, the media, and the general public, who frequently challenge the scientific legitimacy of our discipline. Formulating a coherent answer provides us a means not only to respond to our discipline's critics (at least on a principled, intellectual level), but also to potentially reshape and improve the practice of political science itself.

One common understanding of scientific knowledge is that it develops from proposing and testing theories. That is, science is defined chiefly by its commitment to subject explanations and hypotheses to empirical tests. This is the view of science summarized by Richard Feynman in the epigraph. Against this conventional wisdom, Clarke and Primo argue in A Model Discipline that political scientists should adopt the position that the central feature of science is instead its use of models. In their view, testing theoretical models is also unnecessary and even illogical. Their book is a forceful, extended argument in support of these claims.

While the book is not aimed directly at experimentalists-indeed, they explicitly avoid any extended discussion of how experimentation relates to their model-based view-and is perhaps most obviously of interest to experimentalists in the game theoretic tradition, I expect that experimentalists of all stripes will be interested in much of what Clarke and Primo have to say. Their work will surely be incorporated into first-year graduate scope and methods or research design syllabi and will influence the way that new generations of political scientists think about their research. I also expect that their book will generate substantial discussion as scholars react to their provocative thesis and this dialogue plays out, directly and indirectly, in the pages of our discipline's leading journals.

Clarke and Primo set out to articulate a clear framework for thinking about what models are and for understanding the many roles that they play. On this front, they are entirely successful. Their mantra is that "Models are seen as objects, thus neither true nor false, and are judged by their usefulness for a particular purpose" (p. 1). Their classification scheme divides models into theoretical and empirical

varieties. The purposes of theoretical models (discussed in chapter 4) are to provide foundations, to organize known facts, to explore mechanisms, or to predict (that is, to generate testable implications). The purposes of empirical models (discussed in chapter 5) are to predict, to measure, to characterize, or to test theories. While the book-length argument builds on their Perspectives on Politics article (Clarke and Primo 2007), there is also good deal of new material (notably, the chapter 5 on empirical models and chapter 6 on explanation); their stance against conventional theory testing also intensifies.

Not content with enhancing our understanding of science by providing a thorough definition of what a model is and a useful typology for the variety of intended purposes, Clarke and Primo embark on a more ambitious, radical mission: to break the "spell that theoretical model testing holds over the discipline" (p. 181). In this way, they stake out an alternative methodological position to the EITM movement's response to Green and Shapiro's Pathologies of Rational Choice. Where the EITM movement embraced the charge to test the implications of theoretical models, Clarke and Primo instead contend that doing so is mostly unnecessary and, furthermore, that it makes no sense at all.

The entire book can thus be seen as a sustained effort to undermine the central place that theory-testing has in the way that political scientists approach their research. In chapter 2, they draw from a variety of prominent publications throughout political science to show that the dominant view of science within the discipline is indeed a form of Hypothetico-Deductivism (H-D): propose a theory, deduce hypotheses, and test the hypotheses with data. They discuss the many ways in which H-D is flawed, not just to show that H-D lacks a legitimate claim as the universal foundation for science, but to discredit it entirely as a valid model of the scientific process. In chapter 3, they discuss the analogy of a model as a map and elaborate the key characteristics of models, drawing heavily on developments in the philosophy of science (and the semantic view of theories) to support their contention that models should be treated as objects and are neither true nor false. In chapter 4, they argue that most theoretical models should be evaluated not in terms of their empirical verisimilitude but in terms of the number and importance of the insights they generate. In chapter 5, in discussing the "illogic" of theory-testing, they stress that this is the purpose for which empirical models are least suited because "pairing a theoretical model with an empirical model, regardless of the method of inference, cannot overcome the problems generated by H-D" (p. 117)-doing so amounts to committing the logical fallacy of affirming the consequent and therefore cannot be justified as a logical foundation for scientific practice.

Clarke and Primo are entirely right about the flaws of H-D. It is indisputable that it fails to provide a universal foundation for scientific inquiry, either in its verificationist form (due to the problem of induction) or its nave falsificationist form (due to the problem of auxiliary hypotheses, as embodied in the Duhem-Quine Thesis). Consequently, slavish adherence to H-D as the model of science along with any insistence, in journals and elsewhere, that all good research must adhere to this formula unduly limits what we can discover and understand about the political world. Progress comes in many forms: the identification of empirical regularities may come from the investigation of casual hunches or it may come from subsequent experimental analyses that probe why the data failed to support a strong theoretical hypothesis, new insights may be generated from sparse theoretical models or complex computational models that appear far removed from reality, and new measures or statistical methods

may provide new ways of interpreting old data. Science is a big tent, and Clarke and Primo rightly emphasize the necessity of bringing models to bear on each of these purposes.

But their case against theory testing is overstated. Some form of H-D remains a useful model of science even if it is, like all models, only partially accurate. Like the five-paragraph essay, it is a useful way to teach students about important features of scientific inquiry and it provides a useful structure for many kinds of research activities. The problem occurs when this useful model is anointed as the one, true model.

The basic goal of theory testing also remains justified in their framework, provided that the nature of theory testing is properly understood. If we accept their view that models are neither true nor false, have limited accuracy, and are purpose-relative, then of course it makes no sense to "test" a model to determine whether it is "true" or "false." Clarke and Primo's insistence that this is what is normally meant by "theory testing" is therefore a bit puzzling. But if, in their framework, a model is intended to have limited accuracy, it is natural to ask: how accurate is it for its intended purpose? In other words, theory testing is intended not to establish whether a model is true or false, but the degree to which it corresponds to the features of the real world necessary to explain it.

Theory testing for the purposes of assessing this correspondence is necessary because of our ignorance about the world. Such ignorance is absent in the case of their favorite subway map analogy. The subway map does not need to be tested because its makers already know that it is an accurate representation of the order of subway stops and where different lines meet. But what if this is not known? The more typical case is one where the modeler has limited or partial information. Suppose instead that you are making a subway map and remember the names of only a few stops, but neither their relation to one another nor the stops where different lines meet. You make your best guess as to what the subway map should look like. But you have no way of knowing whether your map is accurate (for its intended purpose) unless you check it against real-world experience (or another map). As you "test" your map, you learn the respects in which your map was accurate and gather new information to determine the ways in which your map needs to be revised. So too with models or theories, which are best guesses about the relevant features of the social and political world that are needed to explain a particular phenomenon, but because they are constructed with partial information must be checked against observation and experiment.

The "theory first, data second" approach just described is one way to develop a theory, but the lack of information with which to construct (even partially) accurate models in the first place suggests that a "data first, theory second" approach can be useful as well. This is essentially what Clarke and Primo advocate in chapter 6. They argue that empirical models can and should be used for characterization and measurement; such models allow us to make better sense of available data and to see a clearer picture of reality, but in and of themselves cannot provide explanations. For Clarke and Primo, the only sensible way to join theoretical and empirical models is to use theoretical models to provide explanations for empirical findings, either in the form of a unifying framework or in explicating a causal mechanism. Here, they explicitly challenge experimentalists to not be content with identifying causal effects.

But no matter how many empirical regularities we might establish and no matter how well a theoretical model seems to provide an explanation, our remaining ingorance still implies an important role for theory testing. Theoretical models must necessarily posit assumptions that cannot be known a priori to be accurate. For example, in game theoretic or rational choice models we cannot completely know whether preferences in a model correspond to agents' preferences in the real world, and we cannot completely know whether real world agents make decisions in the ways prescribed or approximated by such models. Because these unobserved features are often critical to the results of theoretical analyses and because their degree of accuracy remains uncertain, it is necessary to check both assumptions and implications against observation or experiment-that is, to test them.

The "theory first" and "data first" approaches are not incompatible. Choosing between them is unnecessary, as both are vital parts of the scientific research process. We need empirical models to interpret increasingly complex data and to establish empirical regularities, thereby increasing our knowledge of what the world looks like. We can then explain these regularities with theoretical models, but a substantial amount of work must be done to explore and develop the theoretical machinery before it is sufficiently useful to apply it to a specific phenomenon. We must still check theories against new data. Testing the empirical implications of theoretical models will lead to the discovery of new regularities that may or may not fit previous theoretical models. This constant interplay between theory and data is what game theorist Roger Myerson refers to as a "modeling dialoguea process in which theorists and empiricists work together interactively on the difficult task of finding tractable models that capture and clarify the important aspects of real situations" (1992, p. 64, emphasis original). I do not know whether anything resembling the notion of a modeling dialogue is endorsed by philosophers of science, but it seems like an appropriate middle ground (embraced by at least a few political scientists, e.g., Powell 1999, Slantchev 2012) upon which to base an understanding of the relationship between theory and data.

On the whole, Clarke and Primo's call to broaden what political scientists consider "good work" or what we consider to be "scientific" will resonate clearly with a wide variety of scholars. A Model Discipline provides an expansive framework for thinking about the role of models and how we should evaluate them. It makes a compelling case that models are invaluable tools for scientific inquiry and that theoretical models need not be tested to be useful. Even though experimentation does not play directly into their discussion, it would be intriguing to think of an experiment as a kind of model and to consider how this experiment-as-model would fit in their framework. The book raises important epistemological issues regarding the foundations of knowledge and challenges political scientists to think critically about how such foundations shape our work, whether that work is experimental, observational, methodological, or theoretical.

References

Clarke, Kevin and David Primo. 2007. "Modernizing Political Science: A Model-Based Approach." Perspectives on Politics 5(4):741-753

Myerson, Roger B. 1992. "On the Value of Game Theory in Social Science." Rationality and Society 4(1):62-73.

Powell, Robert. 1999. In the Shadow of Power: States and Strategies in International Politics. Princeton University Press

Slantchev, Branislav. 2012. "Audience Cost Theory and Its Audiences." Security Studies 21: 376-382.

Section News and Announcements

- **WCE 2013**

From Kevin Esterling

WCE.2013, Stanford: The sixth annual meeting of the West Coast Experiments Conference will be held at Stanford on Friday, May 10, 2013. We encourage anyone with an interest in experiments to attend; graduate students are especially welcome, as well as those who are new to experimental research. The WCE conference is organized more as a methods "workshop" than as a venue to engage in subfield debates. Instead of standard conference presentations, presenters focus in depth on one or two methodological take away points of their experimental work. The goal is to give the audience members applied, practical advice on methods and design in a way that will help them improve their own experimental research. The WCE meeting is a single day meeting, starting at 9 and ending after dinner. Although we do not have the money to cover travel or lodging expenses, we will provide all meals on that day and we promise a good conversation.

- **Study registry conference**

From Kevin Esterling

Conference at UC Berkeley on Dec. 7, on designing a study registry for social science.
http://cega.berkeley.edu/events/Pre-analysisPlans/

- **Existing Newsletter Sections Need Authors**

I need future authors!! Book reviews (nominate your own book and reviewer too!), theme sections (e.g., the bio-politics and GOTV sections), co-author/inter-disciplinary experiences....Submit! Next issue is May 2013.

- **Call for book proposals**

Call for Proposals: Routledge Studies in Political Psychology www.routledge.com/politics

Series Editors: Howard Lavine (University of Minnesota)

Advisory Board: Ted Brader, University of Michigan; Eugene Borgida, University of Minne-sota; Marc Ross, Bryn Mawr College and Linda Skitka, University of Illinois, Chicago

Political psychology is dedicated to the analysis of the interrelationships between psycho-logical and political processes. The field is interdisciplinary in nature, bidirectional in influence (the psyche influences political orientation and the polity leaves its mark on who we are), and draws on a broad range of disciplinary sources, including cultural anthropology, history, economics, neuroscience, behavioral genetics, sociology, and organizational be-havior. From a psychological perspective, politics is not only about who gets what in the material sense; it is also about how psychological predispositions, social identities and bounded information processing capacities fundamentally shape and constrain how individuals interact with their government and society at large. The Routledge Studies in Political Psychology was developed to publish books representing the widest range of theoretical, methodological and epistemological approaches in political psychology. The series is intended to expand awareness of the creative application of psychological theory within the domain of politics and foster deeper appreciation of the psychological roots of political behavior. We are particularly interested in scholarly mono-graphs, but proposals for edited volumes will also be considered. If you have a book proposal or idea in mind that you would like to discuss or any questions about the series, please do not hesitate to contact:

Series Editor Howard Lavine: lavine@umn.edu

Acquisitions Editor Natalja Mortensen, Political Science Research: natalja.mortensen@taylorandfrancis.com