



The Experimental Political Scientist



In this issue

- [Wilson on publishing](#)
- [Rubenson on Dunning](#)
- [Lab Report 1: Oxford](#)
- [Lab Report 2: Stony Brook](#)
- [Conference Report 1: Voting experiments](#)
- [Conference Report 2: Behavioral politics](#)
- [Interpreting confidence intervals in graphs](#)
- [Announcements](#)

Welcome to the Spring 2014 issue of the *Experimental Political Scientist*. We begin with a discussion on publishing experimental science by Rick Wilson, a long-time editor of the *AJPS*. Daniel Rubenson reviews Thad Dunning's latest book. Ray Duch (Nuffield College) and Matthew Lebo (Stony Brook University) provide dispatches on their experimental labs. Blais and colleagues provide a wrap-up from their Making Electoral Democracy Work workshop on voting experiments and Jonathan Woon does the same from his great behavioural politics conference. The newsletter concludes with a clear and clever piece by Toby Bolsen and Judd Thornton on a common mistake in interpreting graphical presentations.

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.

From the Editor

This is my first edition as editor of the newsletter for the Experimental section of the American Political Science Association. I wish to begin, then, by thanking the previous editor, Dustin Tingley. He not only curated interesting content, he also built up an impressive infrastructure for the publishing the newsletter. So, I am in his debt. I wish also to thank the contributors to this issue. They are productive and otherwise in-demand political scientists, so I appreciate them taking the time to contribute to the newsletter.

Political science is an increasingly experimental discipline, and the content of this newsletter reflects this. Labs are being built to conduct new research. Journals are increasingly receptive to well-presented research. New techniques for inference from natural and quasi-experiments are being disseminated. And interesting conferences are happening. All of this is reflected in this edition of the newsletter. Going forward, I look forward to any suggestions or critiques on the newsletter. Please be in touch and, until then, have a randomly good time.

Publishing Experimental Political Science

Rick Wilson
Rice University
rkw@rice.edu

I have been asked to comment on publishing experiments in general journals. The good news is that general journals increasingly are publishing articles with experiments. This is a far cry from 30 years ago when it was exceedingly difficult to get an experiment into a general journal. Editors are more savvy concerning manuscripts that have an experiment. For example, few editors reject a manuscript because a reviewer complains that there is no external validity. Most editors have become educated to what experiments are capable of showing and which questions can appropriately be answered with an experiment. The bottom line is that there is no hostility toward experimental research among the gatekeepers.

For getting published, the usual caveats for a general journal apply. The work should be excellent. The manuscript should tackle an important question of general interest. It should offer a novel theoretical insight or tie a literature together. Obviously, as an experiment, it should provide a clear result. A great experiment provides an answer. When designed appropriately the experiment should generate an unambiguous answer, either a positive result or a clear null result. Either result should be definitive and leave the reader no doubt about the finding.

While on the subject of a general journal keep in mind several points. The manuscript you send to a general journal ought to be polished and complete. At a general journal you are only going to get one shot. If the editor or reviewers deems the manuscript as a rough draft, it will be dismissed.

As well, know your audience. Is the question you are asking of general interest to political science? You might have a well-crafted experiment that generates clear results. However, if the result is only important for the subfield you should save some time and energy and instead send it to a subfield journal. It is likely this is where the manuscript will finally land. While the incentives are skewed toward publication in general journals, for those considering tenure and promotion, your portfolio will include a wide body of work. All of us read specialized journals and would love to see your work sooner rather than later.

Inevitably most of your work will be rejected by a general journal. No doubt the editor and the reviewers are idiots and cannot see the importance of your work. Get angry, write a diatribe and then shelve it for three days. Then go back to the editor's letter and the reviews. Read what they have to say with an eye toward how you can improve your manuscript. Revise your manuscript before resubmitting it it may well be that one or more of the same reviewers will again be chosen. If you have ignored their advice the first (or second) time, how likely are you to take direction from an editor who asks for a revision? From the editor's standpoint, it seems unlikely.

And now for some specific advice concerning experiments.

1. Avoid overstating what the experiment can do. You may feel the need to make grandiose external

validity claims. Don't. Experiments, when done well, answer very clear and specific questions. Do not feel the need to extrapolate beyond what you can credibly say.

2. A single experiment will probably not suffice. If you are going to address a key theoretical or empirical question you want to be certain that you have addressed the most plausible alternative explanations. This may mean multiple experiments to rule out possible threats to inference. A manuscript that is of general interest will be tackling an important question. A single experiment is unlikely to definitively answer the question.
3. Include everything on submission. Your reviewers will want to see how the experiment was conducted. Submit the full protocol as part of the supporting information. You will want to post this information later anyway.
4. Be prepared to run additional experiments if you get an R&R. Obviously this may not be feasible for a field experiment and most editors (if not reviewers) are sympathetic to this possibility. However, for a laboratory experiment, it may be that an additional treatment is needed. Be prepared to run one. However, at the same time, think creatively about what experiment needs to be run. Often the question raised by the editor or the reviewer does not require a full factorial design. It may be that a single cell needs to be addressed.
5. Replication. This is a tricky question for a general journal. Most are not interested in publishing pure replications. However, if your experiment builds off a canonical experimental design, then it is useful to replicate key features of that design. Such results should be reported; if nothing else, they serve as a baseline for your own experiment and tie your experiment back to broader results. If the experiment is a pure replication it is unlikely to land in a general journal. This is unfortunate since replication is important. Yet there is limited room in the journals.
6. Consider who fair reviewers might be. Suggest them to the Editor. Editors may or may not go along with your suggestions. But if you offer a good reason for why a particular reviewer would be most appropriate you are doing yourself and the Editor a favor. Editors need the best advice they can get. If you direct them to the most appropriate reviewers who understand experimental design and the substantive question, this is all the better.

Following these general points is not going to guarantee publication. After all, the success rate for all manuscripts at general journals hovers about 8 percent. However, there are many things that you can do that will improve your chances. Some are noted above. Elsewhere I have written on the Seven Deadly Sins of publishing. You might check it out at:

<http://ajps.org/2013/12/06/publishing-and-the-seven-deadly-sins>.

The Promises and Pitfalls of Natural Experiments: A Review of Thad Dunning's *Natural Experiments in the Social Sciences: A Design-Based Approach*

Daniel Rubenson
Ryerson University
rubenson@ryerson.ca

The term *natural experiment* is firmly in the social science lexicon. But there is of course a difference between “natural experiment” in common parlance—even when used in academic writing—and “natural experiment” as a term of art. Until recently the place of natural experiments in the social science toolkit was less well established. As widely used (and perhaps misused) as this term is, we have lacked a systematic methodological treatment of natural experiments. Thad Dunning’s book provides exactly this.

Working out the causes of things is one of the most important and challenging tasks social scientists engage in. Over the last decade or so there has been an increase in the interest in and attention to causal inference. The fact that this review is being written for the newsletter of an organized section of the American Political Science Association on experimental research is clearly evidence of this. As is the section’s journal, and of course the proliferation of randomized field experiments, lab and survey experiments as well as new research networks such as Experiments in Governance and Politics (EGAP) and others.

Randomized controlled trials are widely seen as the “gold standard” when it comes to causal inference. But many have noted, including Dunning in *Natural Experiments*, that random assignment is sometimes impractical or impossible and sometimes even undesirable. As a result, numerous scholars are turning to naturally occurring “experiments” with random or as-if random assignment of some cause of interest.

Of course natural experiments come with their own set of challenges. One of the great things about Dunning’s book is that he acknowledges and deals with these challenges head on, providing a set of practical tools that scholars can use. Indeed, discussion of these challenges—in particular how to find natural experiments and, crucially, how to evaluate claims about random or “as-if” random assignment—form large parts of the book as opposed to short afterthoughts.

The book is organized around a number of central questions: How and where can we find natural experiments? What are the most appropriate methods for analyzing natural experiments—and how should we best combine quantitative and qualitative analytical tools? How do we know if a natural experiment is any good? That is, what criteria should we use to evaluate the strengths and shortcomings of particular natural experiments? And how can we build and improve upon research designs using natural experiments?

Dunning devotes one part of the book to each of these broad questions. After setting the stage with a set of clear examples that serve to introduce the approach and provide some basic definitions, Part I of

the book takes readers through the variety of natural experiments that might be exploited in research agendas in the social sciences—from natural experiments with true randomization and as-if randomization to regression discontinuity and instrumental variable designs. Dunning carefully outlines their particular features and challenges.

In Part II Dunning provides a guide for analyzing natural experiments. He marshals the tools of both quantitative and qualitative analysis, beginning with an introduction to the potential outcomes framework and a treatment of the particular considerations we need to attend to when it comes to sampling and standard errors in natural experiments. Dunning then moves to the importance of qualitative methods in the analysis and evaluation of natural experiments. This part of the book is particularly compelling since Dunning makes a serious—and I would suggest successful—attempt to provide a systematic framework for combining quantitative and qualitative methods. This is in contrast to the often superficial treatment of this combination.

The writing in the chapters of Part II is largely nontechnical. However, all of the details of these methods are included in well written appendices pitched at just the right level. The focus throughout is on how to apply these methods correctly to substantive questions of interest in the social sciences.

In the introductory chapter of the book, Dunning develops a three dimensional typology according to which research designs can be evaluated. The elements of this typology are: “(1) the plausibility of as-if random assignment to treatment; (2) the credibility of causal and statistical models; and (3) the substantive relevance of the treatment” (Dunning, 2012, 27). Questions of how to tell the good from the bad when it comes to natural experiments are covered in Part III where Dunning elaborates on this typology, devoting a chapter to each of the three dimensions.

As the subtitle of the book suggests, Dunning’s interest in natural experiments is part of a concern with research *design*. That is, this is a part of the broader recent move toward identification driven approaches to social science where attention to causal processes comes prior to the statistical tools with which to analyze data. One of the great features of the book is that it weaves this attention to design through every aspect of the discussion of natural experiments, from thinking about where to find them to learning how to analyze them and how to evaluate them. In this way, *Natural Experiments*—and in particular Part III and the concluding chapter—makes a contribution to how we ought to think about social science research design more broadly, beyond the distinct topic of natural experiments.

References

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

Lab Report 1: The Centre for Experimental Social Sciences, Nuffield College, Oxford

Raymond Duch
University of Oxford
raymond.duch@nuffield.ox.ac.uk

The Centre for Experimental Social Sciences (CESS) was created by Nuffield College in February, 2008. Its principal goal is to promote and facilitate experimental research by social scientists at the University of Oxford. Nuffield College has a strong commitment to promoting cross-disciplinary scholarship in the social sciences, and the experimental initiative is intended to make a major contribution to this endeavor.

The CESS has experimental lab facilities located right next to the Nuffield College. The CESS lab facilities became operational in October 2008, with twenty-five fully partitioned workstations and a separate experimenter's computer. It is designed to support a wide range of experiments including those in which subjects are entirely separated from other participants. For lab experiments the CESS has two large subject pools of University of Oxford students and non-students, each of which has more than a thousand registered subjects. In addition, the CESS has been developing an online subject panel recruited from all over the United Kingdom. More than five hundred people have signed up for the online panel. The laboratory and online facilities are open for use by both University staff members and other interested parties.

For developing and maintaining a reputation among the population of current and future subjects, under any circumstances the CESS strictly prohibits deception in the lab. Experiments that involve deception and then later truth-telling (i.e. debriefing) are also proscribed. The CESS rule of no deception meets the highest standard applied to experimental economics.

CESS has a very active roster of lab and online experiments. In a typical year, the CESS conducts approximately 90 experimental sessions in the lab that involve about 1,500 subjects. These experiments covered a range of disciplines that tend to be concentrated in Economics, Politics, and Sociology although the range of social science disciplines have been represented including Biology, Zoology, Psychology, Anthropology and Law. The CESS also has been running large-scale online survey experiments. These began in 2009 with the Comparative Cooperative Campaign Analysis Project (CCAP), an initiative dedicated to studying electoral campaigns through online survey experiments. More recently CESS has conducted online experiments with large scale national samples in the U.K., France and Germany.

The experiments at CESS generate a significant number of research publications on a number of academic journals including the top journals of relevant fields. Following is the selected list of academic journals on which a paper from the CESS lab and online experiments are published: *American Economic Review*, *American Journal of Political Science*, *European Sociological Review*, *Social Networks*, *Animal Behavior*, and *Proceedings of the National Academy of Sciences*. Experiments at CESS have been funded by the leading social science funding agencies in the world including the ESRC, the British Academy,

the U.S. National Science Foundation, the Swiss National Science Foundation, the European Research Council, and the Canadian Social Sciences and Humanities Research Council.

CESS has a very strong commitment to advancing experimental social science research and scholarship. Accordingly, the Centre offers courses and workshops in experimental methods: There are CESS experimental methods workshops conducted at Oxford and in North America at the University of Toronto. CESS provides experimental methods training in the Oxford graduate programmes. And the Centre holds periodic workshops on themes in experimental methods such as Experimental Methods and the Law. Over the course of the academic year CESS has regular seminars that feature the world's most prominent scholars in experimental social science. CESS also has a colloquium series that provides young scholars and graduate students and opportunity to showcase their research. As part of its effort to promote the development of experimental social science at Oxford, CESS has a visiting experimentalist programme that funds visiting experimentalist who spend extended periods of time in residence at Nuffield College.

More detailed information about the CESS can be found on our website redesigned in 2013:

<http://cess-web.nuff.ox.ac.uk>

Lab Report 2: Center for Behavioural Political Economy, Stony Brook University

Matthew Lebo

Stony Brook University

matthew.lebo@stonybrook.edu

In 2011 New York State passed SUNY 2020, legislation that allowed five successive years of tuition increases while insisting the new funds would be tied to faculty hiring. Stony Brook University began planning how it would allot 250 anticipated new faculty lines and set the first step as a SBU-wide competition for inter-disciplinary cluster-hire proposals in 2012. Fifty proposals were submitted and among the 5 winners was Behavioral Political Economy (BPE) a program that will consolidate scholars from the Departments of Political Science, Economics, Psychology and the College of Business into a new laboratory and Center. Approval was also given for 5 new faculty lines.

Behavioral Political Economy begins with the individual-level mechanisms, beliefs, and decision-making processes that relate to political and economic decisions, which naturally connect with the rapidly growing field of Behavioral Economics. The field of BPE brings together shared interests in the areas of regulation, public policy, behavioral and experimental economics, consumer theory, behavioral decision theory, and applied statistics/econometrics. In particular, we look to expand upon Stony Brook's strength in political psychology to become a leader in the wider field of experimental political and social science. A new Center of BPE and its two new laboratories, coupled with our existing Center for Survey Research and Laboratory for Experimental Research in Political behavior will soon house a dozen experimental political scientists and another half-dozen experimentalists from around campus.

The CBPE has always been conceived of as an interdisciplinary endeavor, intended to bring together diversely trained researchers who share a common interest in experimental and behavioral social science. The Center is organized so as to maximize the circulation of ideas and the interdisciplinary cross-fertilization of research agendas. Thus the CBPE will host guest speakers, regular seminars and small thematic conferences, and will be a physical and intellectual space in which faculty, post-doctoral fellows, graduate students, undergraduate students and visitors meet and collaborate.

The CBPE will include two experimental laboratories. The first was just completed at the College of Business and a second 2,500 square foot lab will begin construction within the Political Science department in May of 2014, to be completed by August. The new space will include 30 work-stations, a conference room, a large seminar room, and secondary office space for post-docs and faculty members outside of political science. The laboratory will not allow deception in its experiments; a separate, upgraded lab for political psychology experiments will still accommodate deceptive studies. We are currently building our subject pool and expect to run a large number of subjects next year.

Current projects within the cluster include work on climate change mitigation, dishonesty, social preferences, redistribution, alliances, moral and ethical behavior, networks and communication, and financial decision making. We welcome proposed collaborative projects from researchers at other universities that would make good use of the pool and the lab.

But the key to the long-term success of the CBPE will be the quality of the faculty it brings together. The Director, Matthew Lebo, and two Associate Directors, Reuben Kline and Peter DeScioli, spent much of this year recruiting 4 of the new faculty lines with great success: John Barry Ryan (Political Science, currently assistant professor at Florida State University), Andrew Delton (Political Science, currently post-doctoral researcher at UC-Santa Barbara), Stefan Zeisberger (College of Business, currently assistant professor of behavioral finance at the University of Zurich), and Samuele Centorrino (Economics, currently visiting assistant professor at Brown University). These new faculty will join us in the Fall of 2014 and help us fill in the last position in the cluster, likely in the College of Business. Department lines, especially in the College of Business, will also be allocated to add faculty to the cluster. Post-doc positions will hopefully be available for the 2015-16 academic year and beyond.

We look forward to welcoming our new faculty and launching the CBPE in full force in the fall semester. Our website is developing and will soon offer information on our construction, research program, conferences, guest speakers, and more.

Conference Report 1: Montreal Voting Experiment Workshop

André Blais

University of Montreal

andre.blais@umontreal.ca Jean-François Laslier

Paris School of Economics

jean-francois.laslier@ens.fr Karine Van Der Straeten

Toulouse School of Economics

karine.van-der-straeten@tse-fr.eu

Damien Bol

University of Montreal

damien.bol@umontreal.ca

On March 28 and 29, we held at the University of Montreal a two-day workshop specifically dedicated to voting experiments. This workshop, organized within the Making Electoral Democracy Work project (<http://electoraldemocracy.com>), aimed at presenting latest laboratory, field and survey experiments related to elections. With 36 papers presented and about 70 participants from the US, the UK, Canada, Germany, France, Spain, Sweden, and the Netherlands, the workshop was a huge success. In this note, we come across some of the findings highlighted by the participants.

In the first panel, the panelists addressed the question of how the number of candidates and the polarization of these candidates affect citizen's behavior. Renan Levine (University of Toronto) for example organized elections between fictitious candidates located along an ideological spectrum. The treatment consisted in the number of candidates running and their location on the ideological spectrum. His analyses reveal that when subjects are confronted to a three-candidate elections, they tend to choose the centrist candidate even if his or her position along is more extreme to what they favor in a two-candidate election. In turn, the author argues that this type 'compromise effect' might explain why the popular support of the Liberal party in Canada increased when the left-wing New Democratic Party entered the electoral race in the 1960s.

In a second panel, Sarah Brierley (University of California, Los Angeles) and her colleagues presented a paper where they study the effect of international observers on the amount of fraud in countries where this practice is common. Relying on a randomized field experiment during the 2012 presidential election in Ghana, they find that observers significantly reduce overvoting and suspicious turnout at polling stations where they are deployed. However, the overall effect on the country level is mixed as the candidates manage to re-locate fraud in districts where there is no political observers.

In a panel dedicated to methodological innovations, Michael Bruter (London School of Economics) presented some early findings of his research project ECREP. He and his colleague Sarah Harrison (London School of Economics) examine the psychological dimension of elections. As a first step, they organized a mock election for an hypothetical 'President of the European Union' in a realistic voting station. The treatment was the device on which subjects cast their vote (ballot paper, voting machine...). Their findings suggest that the type and intensity of emotions triggered by the election is

largely affected by this treatment. Also, the originality of their design consisted in the way they measure emotion. Instead of relying on self-reported questionnaires, the authors used the filmed shadow of subjects in the voting booth which is revealing in this respect.

In a panel on the impact of information on voting, Kai Ou (New York University) presented results from a laboratory experiment he conducted with Rebecca Morton (New York University) where the payoff structure give an incentive to some of the subjects to free-ride at the expense of the group utility. The treatment was whether the vote was public or secret. As one can expect, the authors find that subjects behave more often in the interest of the group when the vote is public. However, and this is very surprising, their analysis reveals that most of these subjects prefer the procedure when the vote is public, even when it goes against their selfish interest. Their finding thus calls into question the iron law of secret voting.

In this note, we only present a glimpse of the very interesting research pieces that were presented during the Montreal Voting Experiment Workshop. The full list of papers can be found at <http://electoraldemocracy.com/montreal-voting-experiment-workshop-1514>. In addition, it is worth mentioning that the panels were followed by some lively discussions between the presenters and the audience. In particular, the last panel was the occasion for participants to discuss (with quite a lot of intensity!) the merits of psychological theories over classic rational choice predictions. Shaun Bowler (University of California, Riverside) noticed that in most papers (see for example Levine's paper described above), rationality is of little help to understand voting behavior. However, other participants, including John Duffy (University of Pittsburgh) reminded us that in the absence of an alternative theory, the rational choice theory constitutes a fantastic benchmark to start understanding one's choice. Needless to say, experimentalists interested in elections will have to address this question in the future.

Conference Report 2: Pittsburgh Behavioral Models of Politics Conference

Jonathan Woon
University of Pittsburgh
woon@pitt.edu

In October 2013, I hosted the Behavioral Models of Politics Conference at the University of Pittsburgh, which I co-organized with David Siegel (Duke). The conference aimed to bring together scholars who value analytical modeling and are also open to sharing ideas that broaden scholars' theoretical toolkits beyond the standard rational choice, game theoretic paradigm. Of particular interest to the experimental community is that part of our goal was to encourage interactions between experimentalists, theorists, and more traditionally oriented empirical scholars. Generous funding came from the University of Pittsburgh's Dietrich School of Arts and Sciences and from Washington University in St. Louis' Center for New Institutional Social Science.

Over 60 social scientists (faculty and graduate students) participated in the conference, which featured a mix of experimental and theoretical presentations as well as a lively roundtable discussion between David Siegel, Rick Wilson (Rice), and John Patty (WashU). You can find papers and slide from the conference website (<http://polisci.pitt.edu/behavioralmodels>).

On the experimental side, Cheryl Boudreau (UC Davis) presented a test of rationality and spatial preferences in the context of local elections. Sera Linardi (Pitt) talked about her field experiment, with Nita Rudra (Georgetown), in India to investigate whether foreign direct investment affects prosocial behavior towards the poor. Becky Morton (NYU) presented an experiment, conducted with Kai Ou (NYU), on the effects of secret ballot on ethical voting. I presented my joint work on competition and strategic communication.

The theory presentations included talks by Eric Dickson (NYU) on the role of symbols in a behavioral game theoretic model of social identity. William Minozzi (Ohio State) introduced an innovative behavioral model of Deliberation as Interactive Reasoning. Maggie Penn (WashU) presented an interesting model of communication and the structure of group associations within a society. Ignacio Esponda's (NYU) talk featured a novel behavioral retrospective voting equilibrium inspired by experimental and observational analyses.

The conference also featured a very high quality of comments from both local and external discussants, including Torrey Shineman (Pitt), Liz Carlson (Penn State), Brad LeVeck (UCSD), Ian Turner (WashU), Alistair Wilson (Pitt), Dan Butler (Yale), Reuben Kline (Stony Brook), and Salvatore Nunnari (Columbia).

Overall, the conference was a great success! If you missed out on the conference last year, don't worry: David and I are organizing a similar conference to be held at Duke in October 2014.

Overlapping Confidence Intervals and Null Hypothesis Testing

Toby Bolsen
Georgia State University
tbolsen@gsu.edu
Judd R. Thornton
Georgia State University
tobybolsen@gmail.com

Political scientists often rely on visual presentations to illustrate the impact of an experimental intervention. The purpose of this note is to highlight that a formal test of one's hypothesis is preferred to visually examining whether or not 95% confidence levels overlap. Indeed, relying solely on examining point estimates and 95% confidence intervals can result in Type II errors. In other words, this informal method of testing one's hypothesis is in fact more conservative compared to a traditional test

The problem of using overlapping confidence intervals to test if the difference between two quantities is statistically significant is well known in the statistical community (e.g. [Goldstein and Healy, 1995](#)), as well as in other fields ([Browne, 1979](#); [Schenker and Gentleman, 2001](#); [Zou and Donner, 2008](#); [Austin and Hux, 2002](#); [Cumming and Finch, 2005](#)). It is certainly known by many political scientists as well. Interestingly, however, [Belia et al. \(2005\)](#) found in a study of researchers in the fields of psychology, behavioral neuroscience, and medicine that a substantial proportion misinterpreted overlapping confidence intervals. Given the increasing use of graphical displays of point estimates and associated confidence intervals, we believe that drawing attention to this issue is worthwhile.

An Illustrative Example

Following [Schenker and Gentleman \(2001\)](#), we begin with an illustration of the problem by considering the following hypothetical scenario. Imagine we are interested in whether or not support for some policy is dependent upon how the policy is described. Let us suppose that for our control group, which we call Group A, 200 of 250 respondents support the policy, $P_A = 0.8$; and for our experimental group, which we call Group B, 175 of 250 respondents support the policy, $P_B = 0.7$. If we construct 95% confidence intervals for Group A, we have $0.8 \pm 1.96 \times 0.81 - 0.8/250$, which gives us 0.750, 0.850; for Group B we have $0.7 \pm 1.96 \times 0.71 - 0.7/250$, which gives us 0.643, 0.757. The confidence intervals overlap and if this were the only piece of information we considered, then we would conclude there was not enough evidence to reject the null hypothesis of $P_A = P_B$. However, if we calculate a 95% confidence interval for $P_A - P_B$, we would have the following: $0.7 - 0.8 \pm 1.96 \times .0252 + 0.0292$ —where 0.025 and 0.029 are the standard errors for P_A and P_B —resulting in 0.025, 0.175, indicating we reject the null hypothesis.

Our simple example illustrates that relying exclusively on examining 95% confidence intervals can result in Type II errors. Importantly, we note this fact is true not only for our hypothetical example, but also is a general issue when determining if two values are statistically distinguishable from one another.

For more formal treatments of this issue, see [Browne \(1979\)](#); [Goldstein and Healy \(1995\)](#); [Payton, Greenstone and Schenker \(2003\)](#); [Schenker and Gentleman \(2001\)](#). Also see [Zou and Donner \(2008\)](#), who offer a general solution to the problem when the only information one has is point estimates and confidence intervals for two values.

An Empirical Example

We next demonstrate this point by replicating an actual experiment showing that communications highlighting a pro-social norm regarding energy conservation shape individuals' perceptions about the importance of taking action (see [Bolsen, 2013](#)). The design is a 2 X 2 factorial that varies the presence of a communication highlighting a pro-social norm (no norm, pro-social norm) and a communication highlighting the efficacy of individual action (high vs. low impact). We focus on a replication of the hypothesis test that exposure to a pro-social norm will increase the importance individuals attach to collective action in this domain. (For specific details on the experimental stimuli and measures, see [Bolsen \(2013\)](#).)

As a simple test of this expectation, we present the average importance rating from the importance scale for each of the four conditions in [Figure 1](#). Additionally, the right hand side of the figure displays the difference between conditions 1 and 2. Most important for the purposes of our argument is that the confidence intervals overlap for each of the four conditions. In other words, if we were to rely exclusively on the left-hand side of the figure, we would incorrectly conclude that there is not enough evidence to suggest that any of the groups differ from one another. However, as the right-hand side of the figure indicates, we can reject the null hypothesis of $X_1 = X_2$. Further, the null hypothesis of $X_1 = X_2$ can also be rejected. To summarize, were one to rely only on the presence of overlapping confidence intervals, then one would incorrectly fail to reject the null hypothesis, whereas a more appropriate test finds empirical support for the theory.

Some Practical Advice

Examining both simulated and actual data suggests a few pieces of practical advice. First and foremost, test the hypothesis suggested by the theory. Whether this test involves a simple comparison of means, an ANOVA, or regression analysis or one of its variants (e.g., a path model), the appropriate hypothesis test should be the centerpiece of any analysis ([Blalock, 1979](#)). Given conventional confidence levels (e.g., $\alpha = .05$), this test will never involve simply comparing 95% confidence intervals.

Second, displaying the results of the survey graphically is absolutely encouraged. We wish to emphasize that we certainly are not arguing that scholars abandon the use of graphical displays of experimental results. Nor are we arguing against providing measures of uncertainty. Indeed, providing readers with conditional means and a sense of the variability of the estimates (e.g., standard error bars or confidence intervals) can be very informative.

Related to this second point is a third piece of advice: given the usefulness of graphically displaying statistical results, we encourage experimental scholars to do so in a way consistent with their theory and hypotheses. One possibility is a figure similar to [Figure 1](#). Another possibility would be to alter

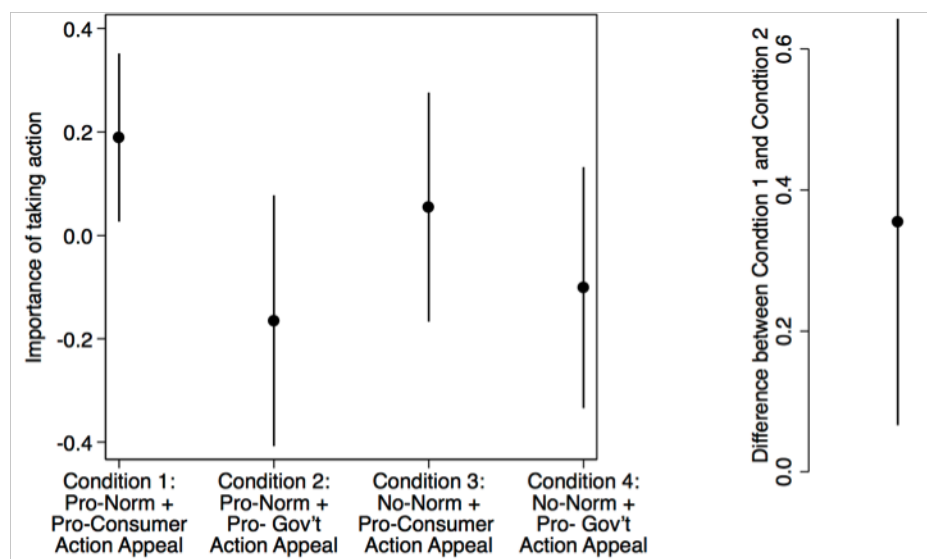


Figure 1: The impact of the experimental conditions on importance ratings. The left-hand side displays the average rating conditioned on experimental condition with 95% confidence intervals. The right-hand side displays the difference between conditions 1 and 2 with 95% confidence intervals.

confidence intervals so they approximately correspond to the results from a formal hypothesis test for example, 83% or 84% confidence intervals for a two-tailed test (Payton, Greenstone and Schenker, 2003). We present such intervals in Figure 2 where they are compared to 95% confidence intervals. Also see Goldstein and Healy (1995) for a more formal discussion of this issue.

Lastly, we should note that much of this advice is equally applicable to those reading and reviewing the research of others. In particular, when evaluating experimental results, keep in mind that relying exclusively on overlapping confidence intervals is overly conservative and inappropriate for evaluating statistical significance. One possibility for readers is to adopt a “rule of eye,” as (Cumming and Finch, 2005, 175-176) suggest. If the proportion of overlap of two 95% confidence intervals is about 0.5, the associated p -value of a two-tailed test is approximately 0.05; and if the intervals do not overlap, $p < 0.01$.

The use of experimental methods has grown considerably over the last two decades. The use of graphs to display experimental results is also increasing. Given these developments, we argue that careful attention must be paid in order to accurately assess if empirical evidence does or does not provide support for one’s theories. We hope drawing attention to the issues raised in this note will aid scholars in the interpretation of their experimental results.

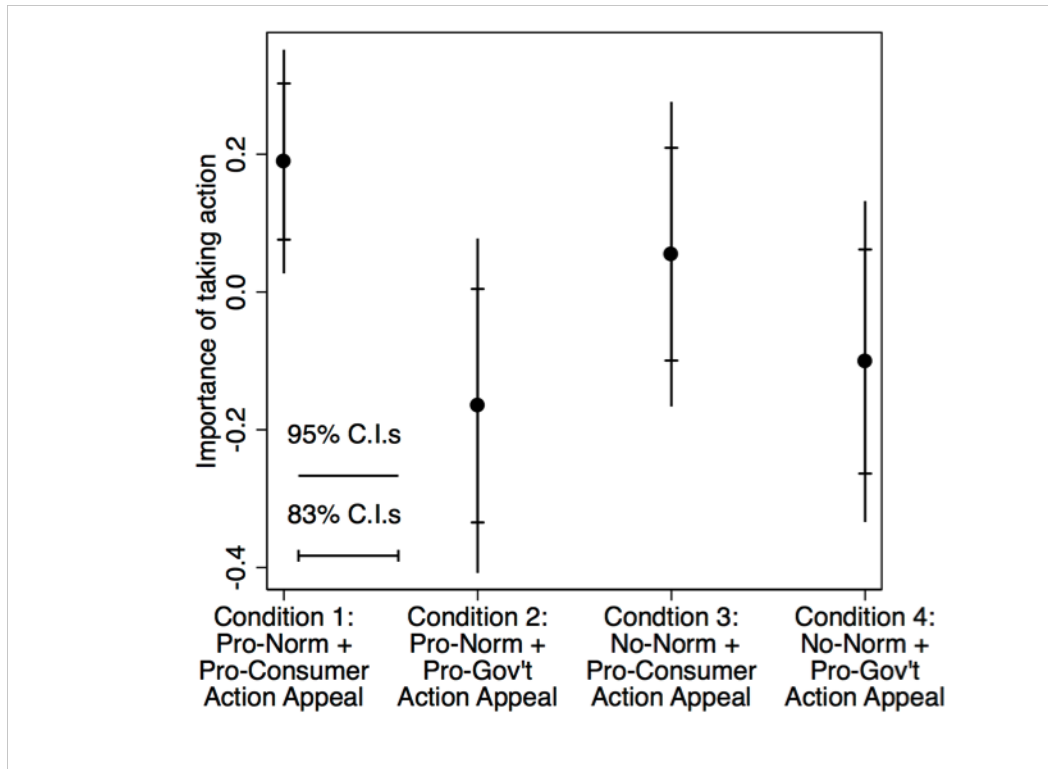


Figure 2: A comparison of 95% and 83% confidence intervals. 83% Confidence intervals are consistent with a two-tailed test when examining overlap with $\alpha = 0.05$ (Payton, Greenstone and Schenker, 2003).

References

- Austin, Peter C and Janet E Hux. 2002. "A brief note on overlapping confidence intervals." *Journal of Vascular Surgery* 36(1):194–195.
- Belia, Sarah, Fiona Fidler, Jennifer Williams and Geoff Cumming. 2005. "Researchers misunderstand confidence intervals and standard error bars." *Psychological methods* 10(4):389.
- Blalock, Hubert M. 1979. "Social Statistics. Revised." New York, NY: McGraw-Hill. Box, GEP, & Cox, DR (1964). *An analysis of transformations. Journal of the Royal Statistical Society, Series B (Methodology)* 26(2):211–252.
- Bolsen, Toby. 2013. "A light bulb goes on: norms, rhetoric, and actions for the public good." *Political Behavior* 35(1):1–20.
- Browne, Richard H. 1979. "On visual assessment of the significance of a mean difference." *Biometrics* pp. 657–665.
- Cumming, Geoff and Sue Finch. 2005. "Inference by Eye: Confidence Intervals and How to Read Pictures of Data." *American Psychologist* 60(2):170.

Goldstein, Harvey and Michael JR Healy. 1995. "The graphical presentation of a collection of means." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* pp. 175–177.

Payton, Mark E, Matthew H Greenstone and Nathaniel Schenker. 2003. "Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance?" *Journal of Insect Science* 3.

Schenker, Nathaniel and Jane F Gentleman. 2001. "On judging the significance of differences by examining the overlap between confidence intervals." *The American Statistician* 55(3):182–186.

Zou, GY and A Donner. 2008. "Construction of confidence limits about effect measures: a general approach." *Statistics in medicine* 27(10):1693–1702.

ANNOUNCEMENTS

- The **Experiments in Governance and Politics** network (EGAP) has issued a request for proposals (RFP) for leading edge experimental research projects that seek to assess the role of information in the selection of effective politicians in developing countries. We invite studies that address the central question using two or more experimental treatment conditions. This is based off of an earlier round that solicited expressions of interest from participants. This grant round is specifically designed to foster knowledge cumulation across studies. Successful applicants will engage in closely related projects and adhere to a common set of research standards. The \$1.8 million pool will support 4-6 research projects that address a common theme and/or one or more grouped applications that link 2-3 individual projects across different research sites. The deadline for submission of proposals is June 16, 2014. Please see <http://e-gap.org/2014/05/03/regranting-1-rfp/> for the RFP, submission forms, and more information about the regranting round.
- The **Toronto Political Behaviour Workshop**, to be held November 7 and 8, 2014, is now calling for papers. This workshop will consist of one and half days of papers. We will also host a seminar on Friday afternoon on text-as-data methods, led by Arthur Spirling, John L. Loeb Associate Professor of the Social Sciences, Harvard University. We will devote one hour to each paper, allowing for a robust discussion of each scholar's work. This will be a wonderful opportunity to get concentrated feedback from a large group of participants. There is no registration fee for the workshop. A dinner will be held on Friday night for participants. Refreshments and lunch will be provided on both days. At this time, we are soliciting proposals for paper presentations. Papers are not at all limited to Canadian politics but should fall under the broad rubric of political behaviour. We are particularly interested in projects that are underway and will be sufficiently advanced to receive useful feedback by the time of the meeting. Please send paper proposals (or drafts) to Daniel Rubenson rubenson@ryerson.ca and Peter Loewen peter.loewen@utoronto.ca. Proposals should be received by June 15 for full consideration. A full call for participation in the workshop will be circulated later in June.

APSA Experimental Section Officers

- President: James Kuklinski
- President-elect: Lynn Vavreck
- At-large Council: Kevin Esterling, Alan Gerber, Jennifer Jerit
- Treasurer: Ted Brader
- Secretary: Neil Malhotra
- Newsletter editor: Peter Loewen
- APSA Program Chair: Betsy Sinclair
- Best Paper at prior APSA: Michael Findley (chair), Jennifer Jerit, Yanna Krupnikov
- Best Dissertation in prior calendar year: Kevin Arceneaux (chair), Cheryl Boudreau, John Bullock
- Best Book in prior calendar year: David Nickerson (chair), Jens Grosser, Jens Hainmueller

[Back to Contents](#)