

L_{aw}

&

C_{ourts}

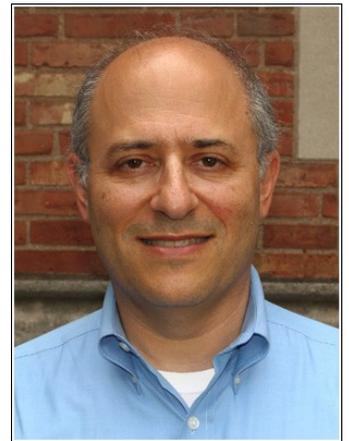
Newsletter
of the Law
& Courts
Section
of the
American
Political
Science
Association

**Volume 24, Number 1
Spring 2014**

Message from the Section Chair: Law and Political Science: Separate Dining Halls?

Gordon Silverstein
Yale Law School
gordon.silverstein@yale.edu

In 1988 Gabriel Almond sadly noted that political scientists were sitting at “separate tables” from one another, divided by ideology and methodology.¹ In thinking about the universe of people who study law and courts – overwhelmingly students and faculty in political science and in law schools – I am struck by the degree to which we not only sit at separate tables, but actually frequent separate dining halls altogether. Sometimes we wander into our counterpart’s citadel – perhaps for a conference, or an individual collaboration; perhaps to find a journal or book in the separate library that is often near the separate dining hall. But in our research and writing, all too often we remain isolated and apart. We are, one would imagine, interested in many of the same questions, motivated by the same impulses, norms and values. And yet the divide is profound.



There are important efforts to bridge this gap: Law professors are increasingly interested in fortifying their work with more sophisticated quantitative methods borrowed from the social sciences.² Indeed, Law & Courts members have played an important role in helping their colleagues in the legal academy to embrace and deploy these methods. For example, our own Lee Epstein and Andrew Martin are about to publish a volume on empirical legal research, dovetailing with the very successful training workshops in quantitative approaches to legal scholarship that they run which enroll a significant number of law professors each year.³

While political science seems eager to teach, to educate, even spread the gospel, the traffic doesn’t run nearly as readily in the other direction. One hopeful sign is another new volume edited by Law & Courts section members Brandon Bartels and

Inside this issue:

Update on the Journal
of Law and
Courts

Symposium:
Experiments in
Law and
Court Research

Tech Notes

Books to Watch For

Announcements:

LSAC Research Grants

Upcoming Conferences

(Continued on page 4)

Table of Contents

<u>Message from the Chair: “Law and Political Science: Separate Dining Halls”</u>	<i>pages 1, 4 – 5</i>
<i>By Gordon Silverstein</i>	
“One Year and Counting: Update on the <i>Journal of Law and Courts</i>	<i>pages 5 – 6</i>
<i>By David Klein</i>	
<u>Symposium: Experiments in Law and Courts Research</u>	
“Opportunities, Issues, and Suggestions”	<i>pages 7 – 9</i>
<i>By Brandon Bartels</i>	
“Some Thoughts on Design and Context”	<i>pages 10–13</i>
<i>By Eileen Braman</i>	
“Utilizing Experiments in Judicial Research”	<i>pages 13– 16</i>
<i>By Brett Curry</i>	
“Recommendations and Reservations”	<i>pages 17– 20</i>
<i>By James Gibson</i>	
“Why Not Experiments?”	<i>pages 21–27</i>
<i>By Justin Wedeking</i>	
Tech Notes	<i>page 27</i>
Books to Watch For	<i>pages 28–29</i>
Announcements	<i>page 30</i>

Officers: Law and Courts Section

<p>Chair: Gordon Silverstein (2013-2014) <i>Yale University</i></p> <p>Chair-Elect: Cornell Clayton (2013-2014) <i>Washington State University</i></p> <p>Treasurer: Chris Bonneau (2012-2014) <i>University of Pittsburgh</i></p> <p>Secretary: Carol Nackenoff (2013-2014) <i>Swarthmore College</i></p>	<p>Executive Committee:</p> <p>George Thomas (2012-2014) <i>Claremont McKenna College</i></p> <p>Matthew Hall (2012-2014) <i>University of Notre Dame</i></p> <p>Anna Law (2013-2015) <i>CUNY Brooklyn College</i></p> <p>Ryan Owens (2013-2015) <i>University of Wisconsin</i></p> <p>Amy Steigerwalt (2013-2015) <i>Georgia State University</i></p>	<p>Law and Politics Book Review Editor Paul Parker <i>Truman State University</i></p> <p>Law and Courts Listserv Moderator Paul Collins <i>University of North Texas</i></p> <p>Webmaster Artemus Ward <i>Northern Illinois University</i></p>
--	---	---

General Information

Law and Courts publishes articles, notes, news items, announcements, commentaries, and features of interest to members of the Law and Courts Section of the APSA. **Law and Courts** publishes three editions a year (Fall, Summer, and Spring). Deadlines for submission of materials are: February 1 (Spring), June 1 (Summer), and October 1 (Fall). Contributions to **Law and Courts** should be sent to the Editor:

Todd Collins, Editor
Law and Courts
Department of Political Science and Public Affairs
Western Carolina University
360 A Stillwell Building
Cullowhee, NC 28723
tcollins@email.wcu.edu

Articles, Notes, and Commentary

We will be glad to consider articles and notes concerning matters of interest to readers of **Law and Courts**. Research findings, teaching innovations, release of original data, or commentary on developments in the field are encouraged.

Footnote and reference style should follow that of the *American Political Science Review*. Please submit your manuscript electronically in MS Word (.doc) or compatible software. Graphics are best submitted as separate files. In addition to bibliography and notes, a listing of website addresses cited in the article with the accompanying page number should be included.

Symposia

Collections of related articles or notes are especially welcome. Please contact the Editor if you have ideas for symposia or if you are interested in editing a collection of common articles. Symposia submissions should follow the guidelines for other manuscripts.

Announcements

Announcements and section news will be included in **Law and Courts**, as well as information regarding upcoming conferences. Organizers of panels are encouraged to inform the Editor so that papers and participants may be reported. Developments in the field such as fellowships, grants, and awards will be announced when possible. Finally, authors should notify **BOOKS TO WATCH FOR EDITOR: Drew Lanier**, of publication of manuscripts or works soon to be completed.

Law and Courts Newsletter

Editor

Todd Collins
Western Carolina University
E-Mail: tcollins@email.wcu.edu

Books to Watch For Editor

Drew Lanier
University of Central Florida
E-Mail: drew.lanier@ucf.edu

Editorial Board

Ryan Black
Michigan State University

Lee Epstein
University of Southern California

Martha Ginn
Georgia Regents University

Tom Hansford
University of California at Merced

Laura Moyer
University of Louisville

Todd Peppers
Roanoke College

Kirk Randazzo
University of South Carolina

Eve Ringsmuth
Oklahoma State University

Michael Tolley
Northeastern University

Claire Wofford
College of Charleston

Chris Bonneau that explores the normative implications of empirical research in law and includes essays by law professors and political science faculty alike.⁴ But here again, the initiative seems to have come from political science (the volume was the product of an NSF grant secured by Bartels and Bonneau).

With a few exceptions (such as constitutional theory, or perhaps legal history, and in some specific areas of study such as equal protection, or election law), law school faculty seem oddly less willing to press their counterparts studying law and courts in political science to employ the methods and approach to the study of law and legal ideas and legal institutions from the perspective of the legal academy, particularly those areas of inquiry in law that do not readily lend themselves to quantitative analysis.

No doubt there are many reasons for this: Traditional legal scholarship appears in student-edited law reviews, a venue that those aspiring for tenure in political science are unequivocally urged to shun. The normative – the “ought” – which is at the very core of legal scholarship is consistently scrubbed out of the leading journals of political science. And from the perspective of the law schools, political science often seems far too narrowly focused, ignoring wide swaths of law, particular private law, and more technical areas such as civil procedure. But another reason may well be that law has not traditionally seen itself as field, or discipline in the way political science, or economics, or sociology have.

Some within the legal academy insist that a discipline is defined by a single, uniform methodology. Lacking that, they argue, law cannot be thought of in those terms. But what social science discipline fully meets that test? Surely not political science where we continue to sit at our separate tables. A discipline, my colleague Robert Post has written, is a “set of practices” by which a body of knowledge “is acquired, confirmed, implemented, preserved and reproduced.”⁵

Law certainly meets that test. Still, we seem to find far less robust debates about what constitutes the

boundaries and content of the field or discipline of law than we find across and among the separate tables of political science. But why?

One reason may be that with rare exceptions, legal scholars do not engage in the same sort of discipline-defining training (and debate) that is commonly a part of the PhD tradition that marks nearly every other field of academic inquiry including political science.

It was, in part, the possibility of helping to spark this sort of conversation within the legal academy and between and among those who study law and courts, that led me to enthusiastically accept an offer from Yale Law School to help establish a new PhD in Law program last year. And though this objective was just one of a number of reasons for this program, I would invite you to consider the role this may play in explaining the distance between these different dining halls, and indulge my briefly dwelling on this particular program.

Law has not traditionally had a PhD in the United States. Many law schools offer a doctoral degree (a JSD or an SJD), but for complicated reasons these degrees are pursued almost entirely by students who earned their first law degree in another country. The legal academy in the U.S. has been staffed primarily by those whose highest degree is a JD, and in some cases, those with a JD and a separate PhD in a related field. But the lack of a PhD degree program may have contributed to the lack of debate about the meaning and contours of the field and discipline of law.

A PhD *in* Law does not mean simply studying doctrine. But it does mean understanding the place and role of doctrine as well as legal reasoning in shaping our legal institutions and behavior. “Judicial decisions may be political, but judicial decision making follows different rules and is driven by different incentives, limited by different constraints, and addressed to different audiences in a different language than is the political process.”⁶ To fully appreciate this requires not only the perspectives offered by economics, political science, history and sociology, but also an understanding of

(Continued on page 5)

law from the perspective of law itself.

A full understanding of the role and impact and possibilities of law also requires that one understand its component parts and how they interact and affect each other: property and torts, contracts and procedure, international law, intellectual property, trusts and estates, criminal law, and so many more. And these interrelationships are often missed or ignored in PhD programs focused on law and courts in cognate fields.

There are important questions in law that cognate fields can and should help answer (and thus there should be little surprise that there is a market in the law schools for what political science has to offer). But just as surely there should be a market in the cognate fields for what law has to teach. Consider how central the rules of civil procedure are, and yet how much has been written on politically-charged constitutional cases and how little has been written about the impact of changes in the language and interpretation of the rules of civil procedure? ⁷ Legal academics should be encouraging their colleagues in political science to explore these aspects of the study of law, but political science itself ought to be reaching out and trying to embrace and incorporate the traditions and approaches of the legal academy to enrich and challenge their own findings and expand the range of questions they ask. Each can cast important light on the other, and each could benefit from considering the questions and

concerns the other has about the world of law and courts.

A small step might involve crossing the campus and sitting down in the other dining hall, figuratively and literally. It is certainly my hope that the Law & Courts Section of APSA can help to facilitate this dialogue not only among our separate tables, but across our separate dining halls as well.

References

1 Gabriel A. Almond, "Separate Tables: Schools and Sects in Political Science." 21 *PS: Political Science and Politics* 4 (Autumn 1988) pp. 828-842.

2 *The Journal of Empirical Legal Studies* which is edited at Cornell Law School, and whose U.S. Editorial Advisory Board includes faculty from ten American law faculties.

3 Lee Epstein and Andrew Martin, *An Introduction to Empirical Legal Research*. Oxford University Press, 2014.

4 Brandon Bartels and Chris Bonneau (editors), *Making Law and Courts Research Relevant: The Normative Implications of Empirical Research*, Routledge (forthcoming). This volume features the work of law faculty and political science alike and might actually be welcomed by Nicholas Kristof who recently lamented the tendency of academics to shun any efforts at policy relevance or impact. (Nicholas Kristof, "Professors, We Need You!" *The New York Times*, February 15, 2014).

5 Robert C. Post, "Debating Disciplinarity" 35 *Critical Inquiry* 749 (2009), p 751.

6 I develop this more fully in Gordon Silverstein, *Law's Allure: How Law Shapes, Constrains, Saves and Kills Politics*. Cambridge University Press (2009), particularly in Chapter 3: "Law Is Different: The Power of Precedent."

7 And here I am myself guilty as charged.

One Year and Counting: Update on the *Journal of Law and Courts*

David E. Klein, Editor
University of Virginia
dek5b@virginia.edu



The *Journal of Law and Courts*, the Section's journal, published its first issue in March, 2013. As we approach the anniversary of that first issue, it seems like a good time to report on where we've come from and where we hope to go. Thank you to Todd Collins for providing this opportunity.

Proposals for a section journal go back some years. The immediate train of events leading to the crea-

tion of it began in late 2008 with the formation of a committee to investigate support for a journal and the feasibility of starting one. The committee's extensive research included a survey of section members, which revealed strong, though not unanimous, support for a new journal.

Perhaps the most urgent recommendation to come from the committee was that the journal "fully reflect the diverse approaches that section members take in their research." The numerous section members who played a role in moving the journal

from idea to reality kept this criterion firmly in mind, and it is reflected in the journal's original call for papers and in the instructions to authors.

Note, though, that there are different ways to accommodate diversity. One is to allocate control over journal space to different disciplinary subgroups. The editorial team is adamantly opposed to such an approach, believing that it would only exacerbate the problems of a discipline that already suffers from intellectual fragmentation. In our view, the most important purpose the journal serves is to build intellectual bridges across substantive and methodological divides, helping scholars to read, learn from, and respond to work very different from their own. We do not ignore the need for intellectual diversity in making editorial decisions, but we place greater emphasis on publishing work that speaks to broad audiences.

Accomplishing our purpose requires more individual and collective effort from board members than at a typical journal. This is reflected in the fact that JLC board members are given the title Associate Editor and in the small size of the board, originally thirteen, now sixteen. In addition to advising on general issues of policy and content, helping line up reviewers, and providing their own feedback on papers, associate editors play a special role. No article is published in JLC until it has been vetted by associate editors who differ from the author in their methodological tendencies or in the part of the world or topics they study. For instance, a quantitative paper on judicial behavior in the United States that is positively received might be sent to one associate editor who studies judicial behavior in other countries and one associate editor who is an expert on the U.S. but focuses on issues other than judicial behavior or primarily employs qualitative methods. It will not be published until those readers agree that the paper succeeds in speaking to a diverse audience.

Of course we have not been able to realize the journal's vision perfectly in just one year. But we are proud of how far it has already come. Counting pieces already published and in the pipeline, there are multiple studies of law and courts outside and inside the U.S., of international law, both trial and appellate courts, and law and society. Qualitative empirical, quantitative empirical, and purely theoretical works are already well represented. We are

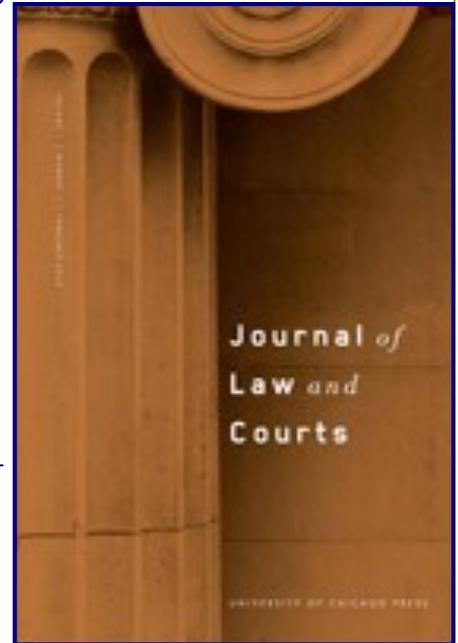
especially pleased that the rate of submissions from scholars outside the U.S. is growing fast. For our part, we will keep pushing hard for greater diversity of scholarship and for authors to engage more actively with other literatures.

What can you do to help the journal succeed? Most importantly, submit your work – your best work – and encourage your students and colleagues, including those from other disciplines, to do the same. Of course we understand that JLC cannot yet have the prestige of more established journals and that prestige of placement

is not a trivial consideration. Still, the most important reason we publish is to share our work with others, allowing us to join and contribute to exciting debates. The vast majority of other available journals will reach only a subset – often a small subset – of readers who might care about and be influenced by your research. JLC goes to every member of the Law and Courts section, and we are aggressively advertising the journal to scholars in other disciplines, including law, sociology, economics, philosophy, psychology, and history.

You can also help with the task of increasing the subscriber base. Please consider submitting a library recommendation form at your institution. We are told by our publisher, The University of Chicago Press, that libraries take recommendations from faculty and students very seriously. You can click on the following link to download the recommendation form for the journal: http://www.press.uchicago.edu/dms/ucp/journals/generaldocs/JLC_Library_Recommendation_Form.pdf.

Thank you for your support, and please don't hesitate to let us know what you think.



Symposium: Experiments in Law and Courts Research

With a renewed interest in experimental research, fueled in part by technological advances that make such projects less burdensome, an outstanding group of law and courts scholars provide useful tips, potential problems, and their overall opinions about experiments in our field.

(Articles appear in alphabetical order by the author's last name,)

Experiments in Law and Courts Research: Opportunities, Issues, and Suggestions

Brandon L. Bartels

George Washington University

bartels@gwu.edu



Political science research on political behavior, public opinion, and political psychology has a rich tradition of employing experiments to test causal effects and causal mechanisms. We have seen a modicum of experimental work in the law and courts realm over the years related to public perceptions of law and courts (e.g., various works by Jeffery Mondak; James Gibson; Rosalee Clawson and Eric Waltenburg) and models of legal reasoning (see Eileen Braman's work on motivated legal reasoning). In this brief article, I offer some considerations for maximizing opportunities associated with experimental design and discuss some issues to watch out for. If used carefully and with the appropriate amount of planning, experimental research has the potential to offer unique and important empirical insights about causality and causal mechanisms that alternative research designs cannot generate. If used carelessly and/or without appropriate planning, experimental research and the results emerging from it will be of minimal value and importance, even if one had noble intentions before embarking on the study.

Before issuing some concrete tips and suggestions, it is worth describing the "state of experimental work in political science" (or at least, my perceptions of that). The types of experiments political scientists conduct are a function of (1) how they are delivered (in the laboratory, via a remote survey, or in the "field" or naturally occurring settings), and (2)

how subjects are recruited (convenience [non-random] samples or probability [random] samples). There are currently four popular types of experimental studies that are employed by political scientists: (1) laboratory experiments using convenience samples (typically college students); (2) survey experiments (most often administered remotely) using nonrandom samples (e.g., Mechanical Turk); (3) survey experiments using random, probability samples (e.g., using a firm such as Knowledge Networks); and (4) field experiments that test the effectiveness of some program or policy (e.g., Alan Gerber and Donald Green's numerous field experiments testing the effectiveness of "get-out-the-vote" strategies on voter turnout). In this essay, I will focus primarily on the first three types of studies, as they have been most relevant to law and courts research.

Consideration 1: Balancing Internal and External Validity

Many discussions about experiments center around the foundational concepts of *internal validity*, or the degree of confidence we have in conclusions about causality, and *external validity*, or the confidence we have in generalizing effects found in the sample to the broader population. The textbook view in thinking about laboratory experiments with college student samples is that they are high in internal validity but low in external validity. Survey experiments using probability samples have been advanced as (partial) solutions to the problem—they substantially boost external validity while still maintaining a high degree of internal validity.

(Continued on page 8)

When it comes to internal validity, laboratory experiments still seem to be the gold standard in that the researcher has sufficient control over what stimuli subjects are exposed to and can minimize confounding and distracting information or content that could interfere with the administration of the experimental treatment(s). With survey experiments administered remotely, the researcher cannot monitor how the study is being administered to the subjects. Are the subjects paying a sufficient amount of attention? Are they reading *espn.com* or Facebook while they are answering survey questions? Are they taking the survey in two or more segments with long breaks in between? Most survey firms report duration measures from the time the respondent started the survey to the time s/he finished. Thus, researchers can filter out respondents with unusually long durations (which indicates that the respondent took the survey in two or more segments). On the whole, this aspect of survey experiments is a threat to internal validity and one that seems more likely to occur in this setting than in laboratory experiments, all things considered.

Survey experiments administered to *nonrandom* samples, e.g., Mechanical Turk studies, are accompanied by both the internal validity problems discussed above and external validity problems as well, since they are not probability samples. Amazon.com's Mechanical Turk (MTurk) is an online labor market where researchers (called "requesters") can pay MTurk "workers" small amounts of money (between \$0.10 to \$1.00 per person per study) to complete a survey (which can include an experimental component) online. MTurk samples are *more* representative of the population than college student samples used in lab experiments, so they are often touted as having an edge in external validity over lab experiments. Though perhaps obvious, it is very important to note that MTurk samples do *not* achieve the level of representativeness that a probability sample achieves. Speaking to the level of external validity attached to MTurk studies, Berinsky, Huber, and Lenz (2012) show that treatment effects using MTurk samples do not differ all that much from treatment effects estimated from probability samples. This study has been used by several researchers to justify the use of MTurk as a powerful, convenient, and relatively cheap (much cheaper than random probability samples) means of conducting experimental work that balances considerations over internal and external validity. On

the other hand, MTurk studies have been criticized on both internal and external validity grounds by scholars such as Dan Kahan.

Consideration 2: Alleviating Validity Concerns with Multiple (and Diversified) Studies

While MTurk offers some advantages in conducting experimental studies (as highlighted by Berinsky et al. 2012), there are enough downsides to maintain a level of concern related to internal and external validity (as pointed out by Kahan and others). I believe that the overall validity of MTurk studies will be a significant target of inquiry in the coming years. But there are, of course, issues (already discussed) with lab experiments and survey experiments with random samples, so the larger question is how to alleviate the inevitable concerns associated with each type of study. One obvious solution, which is sometimes implemented, involves reporting results from *multiple studies*, and preferably, diversifying studies in terms of type of experiment (e.g., lab versus survey experiment). This can help capitalize on different studies' strengths regarding external or internal validity; the strengths of one type of study can partially offset the weaknesses of another. For example, report results from a laboratory experiment, which is higher in internal validity, and a survey experiment, which is higher in external validity, in the same article or book. If the results converge across studies, it presents quite potent evidence for the hypothesis and/or causal mechanism one is testing. Thus, reporting results from multiple studies and working more broadly to pair studies with different approaches can help alleviate concerns about internal and external validity. And of course, a good and fairly common practice is to pair results from an experimental study with results using observational data (e.g., from a national survey employing a probability sample).

Consideration 3: Planning, Pilot Studies, and Research Registration

One cannot say enough about the importance of planning when one is thinking of employing an experimental design. This is all the more important when one is relying on an expensive study, like a probability sample. A researcher often has only one shot at it, so there is pressure to get it right. This underscores the importance of administering test

(Continued on page 9)

runs of an experimental design using pilot data. This issue is relevant for any survey-based study, not just an experimental approach. And for grants and other competitive situations (e.g., Time-sharing Experiments for the Social Sciences, or TESS), funding organizations often want to see results from pilot data to make sure the experimental design is worth funding and implementing. In sum, a researcher should have a very clear idea of what he or she wants to test, which should be guided by good, well-thought-out theory and critical thinking about causal mechanisms. Collecting pilot data first as a sort of “practice run” prevents “woulda-coulda-shoulda,” ex post regret. E.g., “I wish I would’ve included this measure.” “I wish I could’ve added a second experimental factor to test a particular causal mechanism.” “I should’ve tested for this causal mechanism in a different way.” Doing test runs and gathering pilot data allows one to catch and preempt these potential regrets *before* running an official study or studies. I note that MTurk seems like a very promising means of collecting pilot data; it can provide a very valuable setup for administering a survey experiment using a probability sample. Echoing the last paragraph, administering and reporting results from multiple studies also helps, but again, when a researcher has, more or less, one shot at a more expensive study, it is important to get it right the first time. That is not to say that he or she could not pair that study with one or two additional studies that use different approaches.

To implement careful planning, many social scientists are taking “research registration” very seriously. In this approach, the researcher specifies and preregisters on a public website what the study will look like and how he or she will analyze the data *before* one actually collects the data (see, e.g., the Winter 2013 issue of *Political Analysis* for a symposium on research registration). This procedure prevents researchers from manipulating model specification until a hypothesis is confirmed. It also prevents the practice of “p-hacking,” whereby researchers essentially go on a fishing expedition in model specification, selectively reporting results from a model that provided them with a statistically significant ($p < .05$) effect(s) for their independent variable(s) of interest. The importance of research registration in experimental research is underscored in situations where a researcher finds an insignificant effect of a treatment but then

searches for conditional effects of the treatment via numerous models with various interactions. Planning and implementing test runs via pilot studies can prevent such actions by focusing on testing mechanisms and committing to certain conditional effects at the outset.

Conclusion

While experimental research has much to offer, analysts must pay careful attention to these important considerations in order to get the most bang for the buck. I have focused on some general norms and practices that I believe are important in political science for moving experimental research forward. And there are certainly other important considerations I have not discussed in depth here, e.g., the importance of employing experimental research to test for causal mechanisms underlying behavior (see Eileen Braman’s work on legal reasoning; see also Milton Lodge and Charles Taber’s work on motivated reasoning) compared to simply testing treatment effects. Testing causal mechanisms often calls for slightly more complex, multi-factor experiments (two or more experimental factors, or treatments, which are fully crossed). On the other hand, one wants to strike a balance between the appropriate level of complexity called for by the research context and a level of simplicity that makes telling the empirical story more feasible and clear.

In sum, researchers can go a long way toward maximizing the experimental method’s ability to confront and shed light on core theoretical inquiries when they: (1) balance and are constantly mindful of internal versus external validity considerations, (2) report results from multiple (and potentially diversified) experimental studies, and (3) carefully plan studies, using pilot studies and hopefully research registration to commit to a certain type of design and data analysis before collecting data.

References

- 1 Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis* 20(3):351-368.
- 2 See “What’s a Valid Sample? Problem with Mechanical Turk Study Samples, Part 1” and “Fooled Twice, Shame on Who? Problems with Mechanical Turk Study Samples, Part 2.” Available at <http://www.culturalcognition.net>.

Experiments in Law and Courts: Some Thoughts on Design and Context

Eileen Braman
Indiana University
ebraman@indiana.edu



Most law and court scholars have not only, a set of questions that interest them, but also a preferred way of thinking about how to “get at” those questions. In graduate school we learn that our methods should be particularly well-suited to the empirical phenomena we are investigating. We are encouraged to develop a broad methodological “tool box” so we can apply the appropriate tests and analyses on data relevant to our inquiries. Still, methodological preferences can persist – and this is not necessarily a bad thing. Looking at similar questions through different empirical lenses can contribute to a rich understanding of phenomena of interest to law and courts scholars. Personally, I tend to think of empirical research in terms of experimental hypotheses, treatments and conditions.

The approach is most rewarding where theory and experimental methodology lead to hypotheses that can be effectively tested through careful and appropriate design. But of course, experiments have their downside as well. First, not all questions lend themselves to effective operationalization through experiments; those questions that do tend to be about how people *think* about courts, judges or cases. If you are fundamentally interested in the dynamics of Court/Congress relations over the last 30 years, a different method is probably more appropriate. Second, researchers that do experiments have different administrative hurdles to deal with in conducting research. Investigations with human subjects require approval from Internal Review Boards. Different universities and experimental facilities have can have their own rules about the recruitment of participants, “informed consent,” the

use of deception, and the manner in which individuals are “debriefed” following participation in experimental studies. It is important to pay attention to such requirements, even if they seem burdensome. Not only do they involve important ethical considerations for professionals involved in the conduct of experiments, but the availability of research grants, experimental resources, and outlets for publication is often contingent on the demonstration of compliance with applicable rules from the earliest stages of the experimental research process.

Finally, experiments can be difficult to get through the review process because they are highly contextual. Even where researchers are appropriately circumspect about the meaning and generalizability of experimental results, reviewers can be quick to dismiss findings based on particular operationalizations or convenience samples that are not nationally representative. I have found that the best way to deal with such critiques is to do so preemptively, by being up-front about the benefits of experimental inquiry for your particular question, but also candid about the limitations of the approach.

Design and Inquiry

The best experiments have hypothesis and designs that look deceptively simple and straightforward. But make no mistake – really good experiments are hard to design. Many graduate students and professors (especially in psychology) have what I like to call a “well-trained intuition” for experimental design. Some people have more innate skill at the effective isolation of causal mechanisms than others, but as far as I can tell it is an aptitude that can be developed over time. Here are some lessons I have learned.

Don't try to manipulate too much in terms your

(Continued on page 11)

treatments or the number of variables you are testing. One of the most valuable aspects of experimental design is the control it affords. Novice experimentalists can easily ruin that by introducing too many differences across treatments that create confounds which impede the ability to make causal inferences. The most impressive experimental results are those where you can demonstrate large differences between groups based on very small, but meaningful, differences in treatments. If for example, participants evaluate identical court decisions differently depending on whether a judge is identified as “liberal” or “conservative” / “Christian” or “Muslim” / “elected” or “appointed” you have probably discovered something interesting and important by changing a single word across treatment conditions.

You should also limit the number of variables you are manipulating in an experiment. Sometimes this can be hard, especially in a subfield where we are used to statistical models with many control variables. But here you need to be mindful of the other primary benefit of experiments: random assignment. If participants are randomly assigned across conditions you can assume that characteristics that might influence their behavior in your sample are evenly distributed across treatments. Thus, it is not necessary to “control” for the influence of partisanship or ideology on participants’ decisions when you are analyzing differences across groups. You can certainly measure these variables and demonstrate their influence with appropriate statistical tools – or if you have a good theoretical reason to believe priors, like partisanship, will interact with one of your manipulated variables you can include it as a factor to test that hypothesis as well, but there is no need to control for such factors per se.

The issue of manipulating too many variables can be most problematic when a researcher has several factors that she has good theoretical reason to believe will be important in a particular judgment. I was once asked to review a design for a scarce experimental resource with six manipulated variables

that were all interesting and theoretically justifiable concerning perceptions of foreign policy. The problem was six was just too many for a single experiment. As a rule of thumb, it is best not to manipulate more than two or three variables in a single experiment. This is because when analyzing experimental data it is standard practice to investigate the main effects and all interactions between manipulated variables. It is hard to predict how variables will affect each other in an experimental context. Often variables that researchers think will be independent interact with other in complex ways. Contrary to conventional wisdom, too many stars indicating statistical significance can be a bad thing. For one, it is very difficult to interpret the substantive meaning of three and four-way interactions in an experiment. Moreover, where significant main effects for individual variables are embedded in significant higher-level interactions their substantive interpretation is obscured. An experiment with six manipulated variables is likely to lead to uninterpretable mush if there is even one higher-level interaction. Thus, it is best to follow the golden rule of research with experimental design and keep things as simple as possible. If you have more than two or three variables of theoretical interest, you may be looking at conducting more than a single experiment to enable clean causal inferences.

Research in Context

Experiments are contextual. This is at once their main benefit and their biggest shortcoming. It is the researcher’s control over context that gives her the power to make such powerful inferences about cause and effect. If two decisions are identical but for one manipulated factor and we observe a difference between experimental groups we can say that factor made a difference. But context is a double-edged sword. Often critics will question whether or not the decision that is the subject of the experiment is “typical” or representative of all the decisions that could have been tested or whether experimental participants from a convenience sample are wholly representative of a particular population of interest. Far more often than not, the answer to these questions will be “no.” Again, experiments

(Continued on page 12)

are contextual and they must be interpreted in that light. This means that those who conduct experiments must have good justifications for particular operational choices and they must be appropriately modest and circumspect about the meaning of findings given those choices. Findings with respect to one case or legal issue may not translate to another. Moreover, studies should be replicated to demonstrate the durability of findings over distinct legal issues and experimental samples.

It does not mean, however, that all experimental results done on anything but a nationally representative sample of voters or jurists should be summarily dismissed or treated as useless. Indeed, I would argue that where findings of well-designed experiments reveal a causal relationship between hypothesized variables and decision outcomes those results should be taken seriously. Even experiments on much maligned undergraduate samples can reveal something important about how people think about courts and cases (ex. Zink et al. 2009; Ramirez 2008).

Concerns about external validity are strongest where there is some theoretical reason to think that differences between experimental populations and the general public would cause the judgments and cognitive processes investigators are looking at to be different (Druckman and Kam 2011). College students are unquestionably younger and more educated, on average, than the general population (Sears 1986), but these traits do not necessarily mean they think differently about every aspect of law and courts that could be the subject of worthwhile experimental research. Where there is not much theoretical reason to think members of an even a non-representative convenience sample think differently about judges, cases or legal issues concerns about generalizability can be overdrawn. Of course, it is up to the researcher to justify the use of any particular sample, but it is important to remember that findings revealing valid causal relationships are no less “real” because they occur in undergraduate or on-line convenience samples.

There are studies that suggest results from on-line samples from services like Amazon’s M-Turk should not be so quickly dismissed (Beinski et al, 2012).

Those findings can be interesting, important and lead to further testable propositions (Kinder and Palfrey, 1993) that merit publication and dissemination to other scholars interested in how individuals think about law and courts. Research is an iterative process. Often experiments are conducted to demonstrate the viability of particular cognitive processes. Showing that thinking *can occur* in a manner suggested by theory can be an important first step in demonstrating that it does occur that way in more sophisticated (or representative) populations of decision makers. Truly representative national samples for experimental research can be expensive and take time to access. Surely, knowledge about individual cognitive processes gleaned along the way can be of significant value.

Moreover, access to more representative experimental participants can depend on one’s ability to design and cogently analyze findings from such samples. The availability of resources like Timed Experiments in Social Sciences (TESS) is an important development for academics with interesting experimental research agendas seeking representative samples on which to test their hypotheses. To successfully gain access to this resource, researchers must clearly justify their experimental design in a proposal that is usually accompanied by findings from convenience samples to demonstrate the likelihood of achieving significant and substantively interesting results.

In short, experiments can be hard to design, require unique administrative hurdles, and often involve iterative administrations (especially if one is interested in achieving the “gold standard” of a nationally representative sample) but law and courts scholars should not be deterred. Experimental research can be extremely satisfying because as it is

(Continued on page 13)

scholars should not be deterred. Experimental research can be extremely satisfying, as it is uniquely suited to reveal factors and considerations relevant in how people think about courts, cases and even legal doctrine. Moreover, as long as researchers are candid about the benefits and limits of their findings in the context of particular designs and experimental samples there is much to be gained at each stage of the process.

References

- Berinski, Adam J, Gregory A. Huber and Gabriel S. Lenz. 2012 Evaluating On-Line Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*. 20(3)351-368.
- Druckman, James & and Cindy Kam. 2011. Students as Experimental Participants in *Cambridge Handbook of Experimental Political Science* (eds) Jamie Druckman, Donald Green, James Klukinski and Arthur Lupia. New York: Cambridge University Press.
- Kinder, Donald and Thomas R. Palfrey. 1993. "On Behalf of an Experimental Political Science." In *Experimental Foundations of Political Science*, (eds) Donald Kinder and Thomas R. Palfrey. Ann Arbor: University of Michigan Press.
- Ramirez, Mark D. 2008. "Procedural Perceptions and Support for the U.S. Supreme Court." *Political Psychology* 29(5): 675-98.

Utilizing Experiments in Judicial Research

Brett Curry

Georgia Southern University
bcurry@georgiasouthern.edu



Experiments have the potential to make important contributions to the accumulation of knowledge in our subfield. To be more precise, experiments are poised to *continue* making important contributions to our understanding of courts and the law on a number of fronts. Yet, it is my sense that substantial variation continues to exist in terms of the degree to which judicial scholars are conversant in the essentials of experimental research. My primary goal here is to provide an overview of experimentation by discussing some of its advantages and its limitations, the ways in which judicial researchers have employed this methodology in the past, and how they may do so in the future. As I see it, experimentation has much to offer. However, in determining the suitability of experimentation to particular research questions, it is important that we recognize its potential strengths while remaining cognizant of its shortcomings.

The Advantages of Experiments

The principal comparative advantage of experiments over other methodologies is, of course, their unrivaled ability to isolate causal relationships. As Braman and Nelson (2007, 944) put it,

“Experiments offer unmatched power for isolating causal mechanisms and making valid inferences about their effects on observed behavior” (see also Druckman, Green, Kuklinski and Lupia 2006; Kinder and Palfrey 1993). Experimentation is, thus, ideal for decomposing complex phenomena, because it allows scholars to distinguish “features that are merely present from those that are essential to the phenomena of interest” (Brody and Brownstein 1975, 222; McGraw and Hoekstra 1994).

This advantage of experimentation can be leveraged by judicial scholars seeking to disentangle the potential causal influence of multiple factors—something that would be impossible in a non-experimental analysis of a single real-world instance. As but one example, Zink, Spriggs, and Scott (2009) investigate the effect of majority coalition size and the Supreme Court’s treatment of precedent on the propensity of individuals to agree with and accept a given decision. By employing an experimental design, these authors can isolate how these different elements impact support for the Court and its decisions. The utilization of experimentation in designs such as theirs enables school

(Continued on page 14)

ars to engage in “analytic decomposition” (Kinder and Palfrey 1993, 15); other key attributes of experimentation, random assignment (but see Gibson, Caldeira, and Spence 2002) and control over the stimuli participants encounter, operate to ensure that other threats to internal validity are kept in check (Braman 2006; 310-311; McDermott 2002).

The Limitations of Experiments

Like any methodology, experiments have their limitations. The most prominent of those limitations revolves around the difficulty in making broader generalizations from experimental results. At the same time, a number of studies have sought to limit these threats to external validity while preserving the ability of experiments to illuminate causal processes (Gibson and Caldeira 2012; Gibson 2009, 30; Gibson, Caldeira, and Spence 2005). For example, Bartels and Johnson’s (2013) work on the ideological foundations of the Supreme Court’s legitimacy embeds experimental manipulations within the context of a national survey. Their research design effectively marries the advantages of survey research with the capacity of experiments to tap issues of causation.

More specifically, their work illustrates the utility of relying on a multi-method approach in addressing important questions in our field. By employing an experiment to validate findings from a national survey, the strength of their conclusions is enhanced significantly. Though certainly not the only example in our subfield one could point to, Bartels and Johnson’s (2013) work is an excellent example of using experimentalism as a supplement to, rather than an exercise in isolation from, other methodologies. It also underscores that “[E]xternal validity is established over time...and occurs through replication” across a variety of contexts and, not infrequently, by employing multiple research methodologies (McDermott 2002, 335; McDermott 2014). These points are especially important because, as Gibson, Caldeira, and Spence have pointed out (2005, 371), it is unusual for an experiment to resolve causal puzzles in their entirety.

Although technological advances have operated to enhance the similarity of many experiments to real-life situations, experiments are almost always characterized by some degree of artificiality (Gaines, Kuklinski, and Quirk 2007). For one, experimental effects can be highly contingent upon the ways that key stimuli are operationalized. If experimental stimuli are too understated, the researcher risks them escaping the notice of participants (Zink, Spriggs, and Scott 2009, 912). Indeed, some enrichment can be appropriate to “ensure that the experimental stimuli produce the intended effect” (Aronson, Wilson, and Brewer 1998, 117). The construction of control conditions can be thorny as well, particularly in instances when the material in those conditions might inadvertently activate potential thoughts or feelings among some participants (see Gibson et al. 2011, 554).

While experiments are valuable tools for isolating cause and effect relationships, capturing the duration of those effects is not something the typical experiment is constructed to address. This is often of critical importance, as is the fact that many objects of experimental investigation do not occur at a single, discrete point in time (Gaines, Kuklinski, and Quirk 2007, 5-7; Gibson et al. 2011). This, I think, only underscores the necessity of pursuing experimentation in conjunction with a variety of other methodological tools.

Finally, there is the possibility of experimental designs becoming intertwined with the real world. This, as has been noted elsewhere (e.g., Gaines, Kuklinski, and Quirk 2007), can create the potential for contamination of experimental results. It also represents a difficulty that can be inherent when scholars attempt to inject experimental treatments with verisimilitude—enhancing the plausibility or authenticity of a given treatment may also increase the potential for such crossover effects to occur, at least in certain situations. Here again, a multi-

(Continued on page 15)

method approach is likely to be the best way to counter such concerns.

Looking Back and Looking Forward

As I noted earlier, law and courts scholarship has already benefitted significantly from experimentalism. That has been particularly true of investigations into the determinants of judicial legitimacy and public opinion about the courts generally. While many experimental studies of these issues have relied on undergraduate participants, technological innovation is facilitating the ability of scholars to rely on more representative pools of adult participants. Such studies have addressed the potential influence of issues ranging from judicial campaigns (e.g., Gibson and Caldeira 2012; Gibson et al. 2011; Gibson 2009), descriptive racial representation (Scherer and Curry 2010), and a variety of factors pertaining to the Supreme Court and its decisions (Bartels and Johnson 2013; Gibson, Caldeira, and Spence 2005) on perceptions of judicial legitimacy.

A second major avenue of experimental research in our subfield has been geared toward assessing aspects of cognition in legal decision making. According to Braman (2006, 311), experiments have the potential to move us toward “a more sophisticated understanding of the ‘conditions under which’ attitudes are likely to impact legal decision making.” Her work (Braman 2012; Braman 2009; Braman and Nelson 2007; Braman 2006) is at the vanguard of this research, both in terms of its elevation of cognition and psychological processes as well as the role of experimentalism in divining the influence of those processes as they may relate to legal decision making.

The broader subfield is paying increasing heed to the relevance of psychology to judicial decision making (e.g., Klein and Mitchell 2010), and I expect experimentalism to be an important part of that discussion. Schauer (2010) asks whether there is, in fact, a “psychology of judging,” and it will be incumbent on scholars in our subfield to probe for an-

swers to that question. To do so, we will almost certainly need to utilize experimentation (see Bartels 2010). Experimental approaches are particularly likely to accompany this task, given their long association with psychology. As scholars of judicial decision making increasingly speak in psychological terms, experiments can help facilitate these cross-disciplinary conversations (Kinder and Palfrey 1993, 16-19).

Scholars in our subfield have recently discussed the need to recognize and embrace the complexity of judicial decision making (see Braman 2012; Braman 2010; Baum 2010; Collins 2008). As I mentioned at the outset, experiments are especially useful in decomposing such complexity. Bartels (2010, 54-55) notes that, in a perfect world, experiments could potentially shed light on the “black box” of judicial reasoning. Unfortunately, the problem of gaining access to judges for use in such potential experiments is a significant one, and we cannot simply presume that cognitive processes operate in judges as they may in others (see Schauer 2010). That difficulty is unlikely to be overcome with respect to the most elite judges in the American system. However, more realistic avenues for research may exist in the states, amongst retired judges (Bartels 2010), or elsewhere (see Guthrie et al. 2001).

Conclusion

To come full circle, I believe our subfield will continue to benefit from experimental studies. As those of us in the Section read, review and critique, or perhaps even undertake those future studies, we must be mindful of what experiments are and are not well-suited to achieve. We must also remember that experiments are rarely ends unto themselves but, rather, represent gradual steps in the steady accumulation of knowledge about particular phenomena. We should aspire to supplement them with additional experiments in additional contexts, and remember the importance of replication and multi-method research designs. But, in light of the

(Continued on page 16)

field's increasing emphasis on complexity and attention to cognitive processes, experiments may have much to offer us in the years ahead.

REFERENCES

- Aronson, Elliot, Timothy D. Wilson, and Marilynn B. Brewer. 1998. "Experimentation in Social Psychology." In *Handbook of Social Psychology* (4th ed.), ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. New York: Oxford University Press, 99-142.
- Bartels, Brandon L. 2010. "Top-Down and Bottom-Up Models of Judicial Reasoning," in David Klein and Gregory Mitchell, eds. *The Psychology of Judicial Decision Making*. Oxford: Oxford University Press, pp. 41-55.
- Bartels, Brandon L. and Christopher D. Johnston. 2013. "On the Ideological Foundations of Supreme Court Legitimacy in the American Public." *American Journal of Political Science* 57: 184-199.
- Baum, Lawrence. 2010. "Motivation and Judicial Behavior: Expanding the Scope of Inquiry," in David Klein and Gregory Mitchell, eds. *The Psychology of Judicial Decision Making*. Oxford: Oxford University Press, pp. 3-25.
- Braman, Eileen. 2012. "Embracing Complexity in Law and Courts: A Psychological Approach." *Law & Courts* (newsletter of APSA Law and Courts Section) 22(3): 18-23.
- Braman, Eileen. 2010. "Searching for Constraint in Legal Decision Making," in David Klein and Gregory Mitchell, eds. *The Psychology of Judicial Decision Making*. Oxford: Oxford University Press, pp. 203-217.
- Braman, Eileen. 2009. *Law, Politics, and Perception: How Policy Preferences Influence Legal Reasoning*. Charlottesville: University of Virginia Press.
- Braman, Eileen. 2006. "Reasoning on the Threshold: Testing the Separability of Preferences in Legal Decision Making." *Journal of Politics* 68: 308-321.
- Braman, Eileen and Thomas E. Nelson. 2007. "Mechanism of Motivated Reasoning? Analogical Perception in Discrimination Disputes." *American Journal of Political Science* 51: 940-956.
- Brody, Richard A. and Charles N. Brownstein. 1975. "Experimentation and Simulation." Vol. 7, pp. 211-263 in *Handbook of Political Science*, ed. Fred I. Greenstein and Nelson W. Polsby. Reading, MA: Addison-Wesley.
- Collins, Paul M., Jr. 2008. "The Consistency of Judicial Choice." *Journal of Politics* 70: 861-873.
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100: 627-635.
- Gaines, Brian J., James H. Kuklinski, and Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15: 1-20.
- Gibson, James L. 2009. *Electing Judges: The Surprising Effects of Campaigning on Judicial Legitimacy*. Chicago: University of Chicago Press.
- Gibson, James L. and Gregory A. Caldeira. 2012. "Campaign Support, Conflicts of Interest, and Judicial Impartiality: Can Recusals Rescue the Legitimacy of Courts?" *Journal of Politics* 74: 18-34.
- Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2002. "The Role of Theory in Experimental Design: Experiments Without Randomization." *Political Analysis* 10: 362-375.
- Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2005. "Why Do People Accept Public Policies They Oppose? Testing Legitimacy Theory with a Survey-Based Experiment." *Political Research Quarterly* 58: 187-201.
- Gibson, James L., Jeffrey A. Gottfried, Michael X. Delli Carpini, and Kathleen Hall Jamieson. 2011. "The Effects of Judicial Campaign Activity on the Legitimacy of Courts: A Survey-Based Experiment." *Political Research Quarterly* 64: 545-558.
- Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. 2001. "Inside the Judicial Mind." *Cornell Law Review* 86: 777-830.
- Kinder, Donald R. and Thomas R. Pfaffrey, eds. 1993. *Experimental Foundations of Political Science*. Ann Arbor: University of Michigan Press.
- Klein, David and Gregory Mitchell (eds). 2010. *The Psychology of Judicial Decision Making*. New York: Oxford University Press.
- McDermott, Rose. 2014. "Research Transparency and Data Archiving for Experiments." *PS: Political Science & Politics* 47: 67-71.
- McDermott, Rose. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10: 325-342.
- McGraw, Kathleen M. and Valerie Hoekstra. 1994. "Experimentation in Political Science: Historical Trends and Future Directions." In Michael X. Delli Carpini, Leonie Huddy, and Robert Y. Shapiro (eds.), *Research in Micropolitics: New Directions in Political Psychology*, (Vol. 4, pp. 3-29). Greenwich, CT: JAI Press.
- Schauer, Frederick. 2010. "Is there a Psychology of Judging?" in David Klein and Gregory Mitchell, eds. *The Psychology of Judicial Decision Making*. Oxford: Oxford University Press, pp. 103-120.
- Scherer, Nancy and Brett Curry. 2010. "Does Descriptive Race Representation Enhance Institutional Legitimacy? The Case of the U.S. Courts." *Journal of Politics* 72: 90-104.
- Zink, James R., James F. Spriggs II, and John T. Scott. 2009. "Courting the Public: The Influence of Decision Attributes on Individuals' Views of Court Opinions." *Journal of Politics* 71: 909-925.

On Experiments in Law and Politics Research: Recommendations and Reservations

James L. Gibson

Washington University

jgibson@wustl.edu



It is indisputable that a fever for experiments has swept the social sciences, and law and politics research has not been immune to the trend. ¹ This movement reflects a thirst for greater certainty in causal inferences, which in turn is a function of advances in theory

and modeling. This is not the first flirtation political science has had with experiments ², but the prescient bettors are betting that the methodology is here to stay.

I have been involved with experiments in my research for nearly 30 years now. All students of survey research know something about basic experimentation through question-wording experiments that have been around forever (e.g., “forbid” versus “not allow”). Though primitive, these experiments meet all of the requirements of experimental methodology.

In addition, I have conducted two additional styles of experimentation. The first involves experimental vignettes, a method in which short stories are read to respondents, followed by questions about the stories. I first learned about this research technique from Joe Sanders (University of Houston Law), who used vignettes in his cross-cultural studies of punitiveness (e.g., Hamilton and Sanders 1992). A second style of experiment involves counter-arguments and persuasion, often in the context of research on political intolerance. This method would not qualify as a true experiment for many inasmuch as the respondents are not randomly assigned to the treatment condition (see Gibson, Caldeira, and Spence 2002). I will focus in this article primarily on experimental vignettes.

Having conducted dozens of experiment and quasi-experiments over the years has led me to some conclusions about “best practices,” or at least “pretty-good-if-not-perfect practices.” Consequently, I may be able to offer some useful advice for those thinking about using the methodology. As will become clear, I see many limitations to current uses of experimental methods in the study of law and politics, limitations that must be resolved if the research methodology is to fulfill its grandiose promises.

I. The Big Picture Research Question

Experiments often (almost inherently) address small research questions. This assertion may surprise some so I will try to make clear what I see as the most useful role of experiments in law and politics research.

No dependent variable within our subfield can be fully analyzed via experiments. Indeed, the beauty of experiments is that models of the dependent variable need not be fully specified; with random assignment to treatment condition, the “all other things being equal” condition is satisfied. Attached to this is a very large number of wonderful statistical properties (e.g., orthogonal independent variables).

There is, however, a large potential disadvantage of experiments: the specific research question the experiment is designed to address may be so small that it seems to many (i.e., reviewers, the only ones who count) that the question is trivial – or, that it belongs in field journal rather than a general journal. I see many experimental papers these days that seem to have skipped the step of painting the

(Continued on page 18)

big picture first.³

Experiments are most valuable when they are embedded in a large and complete causal model, even when few linkages in the model are susceptible to experimental analysis. It may well be that, of the many individual linkages in a model, only some are amenable to testing via experiments. Social identity theory provides a great example of a theory that actually requires a number of discrete linkages to be activated to get from the exogenous construct (ingroup identification and sympathy) to the endogenous construct (repression of or violence toward the outgroup – see Gibson 2006). Experiments can be helpful in establishing causal connections for some of the linkages in the overall model, even if not all of the linkages in the overall model.

So: my first recommendation is that the experimental research design be firmly grounded in a large and comprehensive model of the dependent variable. Doing so allows a quick and easy answer to the charge that the experiment is trivial. We need to be told why a specific small research question is important; embedding that question in a larger, more comprehensive theory provides the best reply.

II. Experimental Manipulations

Devising a simple experimental manipulation can be pretty simple: subjects can be randomly assigned to read different versions of a story or a concocted newspaper article, for example. For some purposes this works reasonably well.

What is often lost in this sort of research design is any sort of verisimilitude. Not all subjects read newspaper these days; not all Americans read with equal ability; etc. Some experimental treatments I have seen are quite complex and require considerable time and effort for respondents to read and grasp. Some are super-simplistic. Many treatments are easy to administer but may well lack any mundane realism to the subjects.⁴

One of the most glaring shortcomings of current experimental research is the lack of creativity in the types of experimental stimuli used. Even the use of video seems uncommon; more complicated stimuli (e.g., a video game) is almost unheard of in our sub-field. Political economists, of course, often have subjects playing games, but one wonders if this is really an advance inasmuch as these games may not seem very realistic to the research subjects. Developing more realistic and engaging experimental manipulations is certainly one way in which experimentation might be advanced.

Finally, it must be remembered that experimental stimuli are always administered in an unreal context. In reality, people do not sit down and give their undivided attention to news reports about the decisions of the Supreme Court. If watching on tv, the spouse and/or kids are often talking the background (or even in the foreground), the phone sometimes rings, and one might be watching the tv or cooking and reading the newspaper simultaneously. Radio reports are even more susceptible to distortions. The simple conclusion is that the effect in the lab is almost always the maximum possible effect one might observe, and the effect in reality is almost always much smaller than the maximum.⁵

III. Manipulation Checks

Not all experimental research uses manipulation checks, and, among studies using such checks, not all scholars employ the information in the same way. It is axiomatic in experimental research that the stimulus administered may not be the stimulus received. Checking respondents' understanding of the stimulus is therefore essential.

Manipulation checks, of course, may themselves fail.⁶ Still, it is imperative that, except for the simplest of manipulation, some empirical indicator of whether the respondents saw and understood the stimuli employed. This is especially the case with internet surveys, in which many respondents are

(Continued on page 19)

simply trying to complete the interview as quickly as possible to reap their promised reward. For those not experienced at conducting experiments, the results of manipulation checks are often shocking and demoralizing.

Some scholars will discard respondents who fail the manipulation check (e.g., Zink, Spriggs, and Scott 2009), and in some instances, this involves a large proportion of the sample (e.g., Simon and Scurich 2011). Perhaps a better approach is to use the manipulation checks to assess the overall success of the experiment. This can be done by answering a series of questions: (1) Did the experimental manipulation have its intended impact? (2) Was the experimental manipulation accurately perceived? (3) What are the relative impacts of the experimental manipulation and perceptions of the manipulation? (4) Finally, can the perceptions be incorporated into a causal model that addresses all of the above questions, with the experimental manipulation being used as the exogenous variable? Failed experiments can on occasion produce interesting substantive conclusions, but only if some specific indications can be gotten as to why the experiment failed. Including manipulation checks in the study is therefore essential.

IV. Dependent Variables

It is axiomatic that single-item indicators are inferior to multi-item indicators. Thus, the dependent variable for the experiment ought to be measured, usually through the use of multiple questions. A question set allows the empirical examination of validity and reliability, the creation of continuous, interval-level indices – with all the attendant statistical advantages – and virtually always creates a measure with less error variance in it than any of the component parts. ⁷

Care must be taken, however, to include in the dependent variable set an indicator that can be used for illustrative purposes. We seem to be hell-bent as

a discipline on emphasizing the communication of results as being as important as the production of results. Consequently, a simple measure that can be included in a graph (“was the outcome fair?”) is often useful as a means of illustrating the findings of the experiment, while at the same time serving as one of the indicators of the latent construct for analytical purposes.

V. External Validity

This brings me to my most contentious recommendation: Attention must be given to the external validity of the experiment. This advice is currently being swept aside by the predominating self-interest of the scientists who are using convenience and opt-in samples. ⁸

This subject ought to be the focus of an additional symposium on research methods because there are many issues involved. Experimentation does not solve the problem of non-representative samples. Moreover, I urge judicial scholars to have a look at the AAPOR Task Force on opt-in samples (Baker et al. 2010), and AAPOR’s subsequent statement on the use of inferential statistics with convenience samples. ⁹

I fully recognize that, as Sam Stouffer taught us in the 1950s, that there are two aspects of a research methodology: its scientific validity and its legitimacy. These two factors are not necessarily strongly correlated. What I worry about is that self-interest has led us to suspend our scientific judgments about survey samples, and that this tendency has become sufficiently broad that entirely non-scientific samples have achieved scientific legitimacy. This is a big question, but I will submit that: (1) undergraduates at Washington U or any other university are not representative of the American people, (2) convenience samples composed of people with extraordinarily high levels of political knowledge are not representative of the American people, (3) people who take hundreds and hun-

(Continued on page 20)

dreds of surveys are not representative of the American people, and (4) surveys with response rates in the single digits, which is true of all KN and TESS/KN surveys, ought not to be treated as representative of the American people.

I therefore contend that the findings of experiments conducted on samples with no claim to representativeness ought to be treated as extraordinarily tentative findings that demand replication on representative samples. Where such tentative findings ought to be published will continue to be a matter of some debate among members of the law and politics subfield.

VI. Conclusions

One of the most influential experimental political scientists has recently give his own advice to the field of Political Communication: “Enough already with the experiments” (Kinder 2007). That is probably not the advice I want to put before the law and politics community, in part because few would heed it anyway.

Instead, I urge the subfield to pay attention to experiments but at the same time to return to the fundamentals of good social science research. Be concerned about the validity and reliability of one’s

measures, whether the experiment creates such an unnatural context that even its internal validity is threatened, and whether research findings are generalizable to some meaningful population. Ask whether the experiment treatment interacts with pre-existing attributes of the respondents, and, if so, whether the distribution of that attributes is the same in the sample as it is in the population. Think big theory, and embed the inevitably small experimental question within a comprehensive model. Remember that research is a continuous, evolving process, including exploratory and confirmatory work, observational and experimental studies, all not with the goal of publishing an individual paper

but rather of understanding an important political and/or legal phenomenon. Perhaps most important, do not suspend the application of your critical faculties to this methodology just because it makes data collection so inexpensive and easy. Experiments have much to offer, but our subfield will not profit from the accumulation of a wide variety of research findings that are highly dependent upon the sample and the context of individual research projects.

*I appreciate the comments on Michael Nelson on an earlier version of this article.

References

- 1 Citing just a few examples from our best journals should suffice to justify this claim. See, for examples, Bartels and Johnston (2013), Zink, Spriggs, and Scott (2009), and Gibson 2008.
- 2 For a great review of the development of experimentation in political science see Iyengar 2011.
- 3 Of course, multi-investigator studies (e.g., TESS) as valuable as they are, encourage small-scale research by only allowing investigators tiny numbers of survey questions.
- 4 On “mundane realism” or verisimilitude see Aronson et al. 1990, and Druckman and Lupia 2006.
- 5 The research h of Barabas and Jerit (2010, 238) indicates that experiments capture the *maximum possible treatment effect*: “[a]lthough the real world does not look so different as to throw into doubt the validity of survey experiments, there is drop-off in terms of both the size of the treatment effect and the population experiencing those effects” when comparing real-world treatment effects to experimental effects.
- 6 See Gibson 2009 for many painful examples of failed manipulations and manipulation checks.
- 7 This is just one reason why it makes no sense to analyze *both* the variance in the index and the individual components of the index.
- 8 It is perhaps worth noting that the cost of a respondent in the General Social Survey is around \$1,500, while the cost of a Mechanical Turk respondent can be as little as \$.15.
- 9 AAPOR has taken a clear position on the appropriateness of using inferential statistics on opt-in surveys: “The reporting of a margin of sampling error associated with an opt-in or self-identified sample (that is, in a survey or poll where respondents are self-selecting) is misleading . . . AAPOR considers it harmful to include statements about the theoretical statements about theoretical calculation of sampling error in descriptions of [opt-in or self-identified . . . Studies, especially when those statements mislead the reader into thinking that the survey is based on a probability sample of the full target population. The harm comes from the inferences that the margin of sampling error estimates can be interpreted like those of probability sample surveys.” See AAPOR, “Opt-in Surveys and Margin of Error,” at http://www.aapor.org/opt_in_surveys_and_margin_of_error1.htm#.Unv7n5V3unA [Accessed 11/7/2013].

Why Not Experiments?

Justin Wedeking

University of Kentucky

justin.wedeking@uky.edu



There has been a sea change in political science regarding attitudes towards experiments. As Druckman, Green, Kuklinski, and Lupia (2006) argue, experimental research has emerged as an undeniable force in the wider discipline. While the roots of this transformation can be traced back to at least the 1950s and

the behavioral revolution, the emergence of political psychology as a distinct subfield in the 1980s can also be credited with igniting a boom of experimental research (Druckman and Miller 2004). In recent years, the number of articles using some form of experimental method has grown remarkably (Druckman et al. [2006] document this growth in the *American Political Science Review*, which had just one experimental article in the 1950s, but over twenty from 1995-2004).

To gain some perspective on this change, consider the role of experiments in psychology, where non-experimental methods are the exception. Experiments remain their “bread and butter” approach to scientific inquiry. And within political science, experiments have been used to shed new light on the role of law and justice in a variety of subfields. In international relations, for example, experiments are now used to examine international law and public attitudes towards torture (Wallace 2013), as well as public support for conflict (Gartner 2008) and the democratic peace (Tomz and Weeks 2013). In comparative politics, experiments are used to examine commitments to property rights and social justice in Benin (e.g., Duch and Palmer 2004) and also the role courts play in conferring legitimacy in civil liberties disputes in South Africa (Gibson and Gouws 2003). Experiments have also been used in a variety of areas of American politics. For example, experiments have been used to study the link

between racialized code words and support for punitive crime policies (Hurwitz and Peffley 2005), as well as to discern the separate realities that Blacks and Whites inhabit when viewing the criminal justice system (Peffley and Hurwitz 2010). In sum, experiments abound in political science.

This discipline-wide change, long in the making, recently achieved an institutional milestone when a new APSA section devoted to experiments was formed, complete with its own newsletter (*The Experimental Political Scientist*) and journal (*Journal of Experimental Political Science*). The rise of experiments can be attributed to an increased focus and emphasis on causal inference and internal validity by scholars in political science. Many argue that experiments are the gold standard for causal inference, which is why the inaugural president of the section, James Druckman, expressed cautious optimism about the future of experiments in political science by noting that this was not the “Golden Era” of the gold standard (Druckman 2011).

Though I agree that caution must be exercised, it is also important to recognize that experiments are now “institutionalized” in our discipline. As many of us know from studying institutions, this suggests that experiments are not just a fad, but will play an important role in political science research in the years to come.

More to the point, what do such changes mean for *Law & Courts* readers? To start, we should be aware of how developments in our field are tied to broader changes in the discipline. While many are undoubtedly familiar with prominent examples of research using experiments in our field (e.g., Braman [2009] and Gibson [2008] are but two examples of this growing body of work), we should also recognize that experiments are a widely accepted and legitimate form of inquiry throughout the social

(Continued on page 22)

sciences. At the same time, readers should keep in mind that experiments are just *one* tool that still needs to be used with caution, just as we would exercise caution when using other tools. This is not to suggest that experiments should (or will) replace another form of inquiry. Rather, in an era when scholars are determined to provide the clearest and most compelling answers to their research question, “triangulating” on an answer by adding experimental evidence can only strengthen one’s approach.

In the rest of this paper, I address three questions. First, why has our field been so slow to adopt experiments? Second, what is all the fuss about regarding representative subject pools and external validity? Finally, I offer some useful “tips” for those who want to “experiment” with experiments. Why has our field been so slow to adopt experiments?

To be sure, the wider acceptance of experiments within the discipline is a relatively recent phenomenon. When I began my doctoral work in 2002 at the University of Minnesota, experiments were not just accepted but encouraged among students in the political psychology program. Such was not the case in the wider discipline, however. I recall numerous stories of manuscripts being rejected because a reviewer objected to the use of experiments (or students as subjects). I was puzzled by such resistance, given the call to experiments in the wider discipline (Kinder and Palfrey 1993; McDermott 2002) and its renewed focus amongst political methodologists as evidenced by the 2002 special issue of *Political Analysis*. Why are there not more examples of experimental work in our field?

Part of the reason, it seems, stems from the nature of what we, as a field, primarily study. Judicial opinions, legal documents, courts, institutional rules and norms, the behavior of judges, attorneys, and juries, as well as how groups in society are influenced by these actors, which can be described as, “all things legal,” do not easily lend themselves to experimental inquiry. Indeed, it is difficult (though

not impossible) to induce judges to be subjects in an experiment (for an exception that uses trial judges as subjects, see Guthrie, Rachlinski, and Wistrich 2007). And of course, law students are often used to study judicial behavior because of their substantial legal training and knowledge (e.g., Braman and Nelson 2007). But the fact remains, those who study institutions, rules, and the like are often faced with a subject of study that simply does not lend itself easily to experimental inquiry. With that said, however, our field has much to gain by embracing experiments. Unfortunately, we do not see enough experimental work in our field due to concerns about the appropriateness of the subject pool and external validity, which I address below.

Subject Pool Considerations and External Validity

There are two important aspects of external validity to consider. The first is whether the results from the subject pool can be generalized to the larger population. The second is whether the finding would apply to another group or population (e.g., whether a finding from a student sample would apply to an adult or non-student sample). Note the second does not necessarily require that the sample be probability based.

All things considered, the subject pool should be selected based on the goals of one’s research. If the goal is to estimate the “true” population parameter, then experiments should be embedded within a probability based survey (or consider a randomized field experiment; see Gerber and Green [2000]). To be sure, embedding experiments within nationally representative surveys will offer some balance between external and internal validity. However, for the vast majority of researchers, the cost of obtaining space on a survey, which can range from \$50,000 (for approximately five minutes) to over \$80,000 (for approximately 20 minutes), is simply prohibitive. In addition, depending on the experiment one has in mind, the expense of a national probability survey may yield diminishing returns as the number of experimental conditions

(Continued on page 23)

increases because each “cell” of the design is assumed to be a representative, random sample. Embedding a 2 X 2 X 3 experiment, for example, with 12 cells into a large sample of 1,200 respondents, will produce cell sizes of 100 respondents each, with large margins of error. The tradeoff is straightforward - as the number of conditions (or cells) increase, the size of the subsamples gets smaller. At what point does external validity lose its added value?

Keep in mind, a single-minded emphasis on external validity (at the expense of internal validity) may not be worth the additional expense. Specifically, research on public opinion over the last 50 or more years has demonstrated that uncovering “true” population parameters is exceptionally difficult and may even be, at times, an illusory quest. We know this because of the large amount of response instability in most political attitudes over time, because of the way questions are framed, the order in which they are asked, and so on.

Furthermore, it is generally more beneficial (at least initially) for testing theoretical propositions to be confident about the internal validity of one’s findings – i.e., to demonstrate that the finding you are studying exists under conditions where you have *more* control over the multitude of extraneous factors that may mask the process by which attitudes are formed and how they move. This means that you might consider using a subject pool that is not from a probability-based representative sample. There are several possibilities, three popular alternatives include: students, adults recruited from Amazon.com’s Mechanical Turk (MTurk), and adults from a module in the Cooperative Congressional Election Study (CCES). I discuss each of these in turn.

Students are readily accessible and the cost is low. But why are people concerned about student samples? Sears (1986) is widely recognized for drawing attention to what is now colloquially referred to as the “college sophomore” or narrow database problem. While Sears aimed his critique at fellow

psychologists, the crux of his concern is that “college students are likely to have less-crystallized attitudes, less-formulated senses of self, stronger cognitive skills, stronger tendencies to comply with authority, and more unstable peer group relationships” (Sears 1986, 515). Stated differently, students are easy to manipulate. However, in a defense of the “narrow database” Druckman and Kam (2011) point out that Sears’ argument is conceptual, and that student subjects “do *not* intrinsically pose a problem for a study’s external validity” (41, emphasis original).

One study that tested Sears’ argument found evidence that student subjects differed on several criteria compared to military officers when confronted with a counterterrorism scenario, and ultimately concluded that student samples are inappropriate (Mintz, Redd, and Vedlitz 2006). However, as Druckman and Kam (2011) note, Mintz et al.’s conclusion is premature, and their own simulation results suggest that it would have been possible to isolate the “key treatment dynamics [in the student sample] that would be found in the military sample” as long as there was some limited variation on the factors suspected to be the reason for finding the difference (2011, 54). In fact, Druckman and Kam show evidence that the conditions necessary to constrain an experimental inference from using a convenience sample are very limited, essentially “only when the size of an experimental treatment effect depends on a characteristic on which the convenience sample has virtually *no* variance” (2011, 1, emphasis added).

A final point about student samples is worth mentioning. Skeptics of student samples are generally quick to dismiss a set of findings because of the belief that students are easy to manipulate (and not representative of the target population), but fail to acknowledge fully the implication of this belief if they ever encountered a “null” experimental finding from a student sample. In other words, if a skeptic is willing to dismiss a finding because of the “college sophomore” problem, this same line of

(Continued on page 24)

thought would suggest the skeptic should be *more* willing to accept a null finding when student subjects are used (See Druckman and Kam [2011] for more on this). In sum, research from leading experts suggests that the fear of student samples has been blown out of proportion.

To illustrate how concern over student samples travels to our domain, imagine how many people would argue that judges – an elite class of decision makers that are trained and experienced in the law – would not fall prey to the same sorts of decision making errors and biases as students or lay-people. However, as Guthrie, Rachlinski and Wistrich (2000) show, even judges are susceptible to the same sorts of cognitive illusions that non-judges are.

If one is still skeptical of students (or law students) as subjects, we currently have a promising option for non-student samples. Researchers are turning to Amazon.com’s Mechanical Turk (MTurk), an online platform for recruiting and paying subjects to perform various tasks. MTurk is now being used in psychology (Buhrmester, Kwang, and Goslin 2011), behavioral economics (Paolacci, Chandler, and Ipeirotis 2010) and political science (Berinsky, Huber, and Lenz 2012). MTurk’s samples have been shown to be more diverse than other convenience samples, have been used to replicate other experiments and have roughly the same treatment effect as population-based surveys (Berinsky et al. 2012). Perhaps the most attractive feature of MTurk is the relative ease by which subjects are recruited. Scholars in political science, in fact, have already begun to make use of MTurk (e.g., Huber, Hill, and Lenz [2012]; Tomz and Weeks [2013]). In sum, while MTurk samples are not probability based, which limits any population-based inferences that can be made, they are an excellent alternative if one is concerned about student samples.

One popular alternative to probability-based nationally representative surveys that has emerged in recent years is the Cooperative Congressional Election Study (CCES), fielded by YouGov. Respon-

dents opt-in and through the use of a two-stage sample matching technique are matched on a number of demographic and attitudinal items using an accepted approach similar to studies in epidemiology. The procedure creates datasets that are nationally representative (e.g., see Vavreck and Rivers 2008; Ansolabehere and Rivers 2013). CCES data is well established in political science and has appeared in a variety of prominent journals (to list a few: Bonneau and Cann forthcoming; Dancy and Sheagley 2013; Harbridge and Malhotra 2011; Jacobson 2012). Even traditional public opinion scholars admit that YouGov has a good “track record” in forecasting elections (Keeter 2011, 48). In sum, we have many good options for subject pools and I turn now to offer some tips for those wanting to try experiments.

Some tips for those wanting to try experiments

My last section offers some tips of the trade that I have found helpful when implementing experiments. First, if one is concerned about data quality, particularly from students or on-line samples where surveys are self-administered, I recommend using a filter question to determine whether respondents are being attentive (Berinsky et al. 2013). The filter question I have used, courtesy of Nyhan and Reifler (2011), gives subjects a long paragraph to read where midway through the paragraph they are instructed to select two specific colors (pink and green) instead of answering the question at the end that asks them their favorite color. We found that approximately 71% of our student sample followed the instructions perfectly, but importantly, *none* of the results changed when we examined only subjects who correctly followed instructions (Wedeking, Peffley, and Ferrell 2014).

Second, pretest your newly constructed questions. While I generally recommend using established and proven questions, do not go into the field with an untested battery of questions. While pretests still need Institutional Review Board (IRB) approval (which can be slow), you can gain valuable insight

(Continued on page 25)

simply by having a class answer the questionnaire and look for flaws.

Third, prepare for the null result, but do not abandon ship if you receive one. Even if your hypothesis lacks support, you may still learn something from your study if you perform a power analysis and include a manipulation check to verify whether your manipulations were perceived as they were intended. For example, as part of a larger project on the role of cameras in the Supreme Court, my coauthors and I studied whether individuals learned about oral arguments differently by reading transcripts or hearing audio versions (Black, Johnson, and Wedeking 2012). With our manipulation check questions we were able to detect that subjects in our conditions could detect differences in the professionalism displayed by the justices speaking, as well as noticing interruptions and whether multiple justices spoke. However, our results showed that this unprofessional and rude behavior at oral argument did not influence the legitimacy of the Court, which was our dependent variable. Because our sample size was small – 116 subjects in a 2X2 plus control design (5 conditions) – there was concern that our analysis may not detect a small effect. However, through the incorporation of a power analysis, which enables one to detect a difference between treatment *groups if it actually exists*, we were able to generate estimates of how confident we would be if we were to miss an effect size of 1 or 2 or 3 points and so on (on our 40 point scale). Thus, while our design was not well equipped to recover small effects, we were able to be very confident that we did not miss any large drops (6 or more points) in legitimacy.

A final recommendation stems from Gaines, Kukulinski, and Quirk (2007). If all the conditions in your experimental design contain a stimulus, be sure to include a baseline or control condition that does not receive a stimulus (e.g., the 2X2 plus control design discussed above). This has two advantages. First, it will allow you to judge the relevance and implications of the effect size from your stimulus if you can compare it to a baseline that did not receive a

stimulus. In other words, you will be able to contrast an effect to a condition that is a better representation of the status quo than some other experimental manipulation that may have “pushed” the effect in the opposite direction. Second, even if the effects of the treatment conditions are not different, it is still possible to detect a difference from the control. For example, when assessing the relative strength of the “pro” and “con” arguments for having cameras in the U.S. Supreme Court, my coauthors and I discovered that receiving *any* combination of “pro” and “con” argument raised opposition to cameras relative to receiving no arguments (the control condition), even though there was no difference between the subjects who received different combinations of the “pro” and “con” arguments (Black, Johnson, and Wedeking 2013).

Conclusion

As a researcher, I strongly think our field has a great deal to gain by embracing experiments. To be sure, experiments, like any single method, are no panacea. And not everyone will agree that Law and Courts subject matter is within the purview of experiments. However, it is important to keep in mind that such views have been changing in political science. To be sure, we will still have difficulty applying experiments to situations that are inherent to what we study. And it is also important to remember that experiments are difficult for another reason, as Druckman and Lupia (2012) point out, politics also entail disagreements over basic values. The same can be said for “all things legal.”

Why else are experiments important? Druckman et al. (2006) write, for example, in terms of policy innovation, “subfields in political science that are currently viewed as beyond the purview of experimental investigation might have a stock of experimental knowledge a century from now” (2006, 634). This is already happening in other fields across political science. Judgments about whether experiments are appropriate are only partly due to the subject matter that we study. There is also a perception prob-

(Continued on page 26)

lem that restricts us from using alternative designs to answer our most important questions. To end, adapting an insight made by Druckman et al. (2006), it is important to remember that experiments have become a generally accepted and influential form of inquiry, and we need not be concerned that the field of law and courts will be turned into an “experimental field,” but with creativity and courage our field can become one whose contributions are enriched, strengthened, and educated by experimental research.

* I thank Ryan Black and Mark Peffley for comments and suggestions. All views and errors, however, remain mine.

References

- Ansolabehere, Stephen, and Douglas Rivers. 2013. “Cooperative Survey Research.” *Annual Review of Political Science* 16(1): 307-329.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis* 20(3): 351-368.
- Black, Ryan C., Timothy R. Johnson, and Justin Wedeking. 2012. “Does the Medium Matter for Learning about Oral Arguments at the U.S. Supreme Court? A First Step to Answering the Television Dilemma.” Paper presented at the 2012 International Conference on Law and Society (June 5-8, Honolulu, HI).
- Black, Ryan C., Timothy R. Johnson, and Justin Wedeking. 2013. “Assessing Support for Televising Oral Arguments: Examining the Relevance of Framing and Legitimacy.” Paper presented at the 2013 American Political Science Association conference in Chicago, IL, August 29-September 1.
- Bonneau, Chris W., and Damon M. Cann. Forthcoming. “Party identification and Vote Choice in Partisan and Nonpartisan Elections.” *Political Behavior*. Published online October 30, 2013.
- Braman, Eileen. 2009. *Law, Politics & Perception: How Policy Preferences Influence Legal Reasoning*. Charlottesville: University of Virginia Press.
- Braman, Eileen, and Thomas E. Nelson. 2007. “Mechanism of Motivated Reasoning? Analogical Perception in Discrimination Disputes.” *American Journal of Political Science* 51(4): 940-956.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. “Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?” *Perspectives on Psychological Science* 6(1): 3-5.
- Dancey, Logan, and Geoffrey Sheagley. 2013. “Heuristics Behaving Badly: Party Cues and Voter Knowledge.” *American Journal of Political Science* 57(2): 312-325.
- Druckman, James N. 2011. “Letter from the President: A Golden Era for the Gold Standard?” *The Experimental Political Scientist: Newsletter for the APSA Experimental Section*. Volume 2, Issue 1, page 1.
- Druckman, James N., and Cindy D. Kam. 2011. “Students as Experimental Participants: A Defense of the ‘Narrow Data Base’.” In *Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York: Cambridge University Press, 41-57.
- Druckman, James N., and Joanne M. Miller. 2004. “The Political Psychology of Electoral Campaigns: Introduction to the Symposium.” *Political Psychology* 25(4): 501-506.
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. “The Growth and Development of Experimental Research in Political Science.” *American Political Science Review* 100(4): 627-635.
- Druckman, James N., and Arthur Lupia. 2012. “Experimenting with Politics.” *Science* 335: 1177-1179.
- Duch, Raymond M., and Harvey D. Palmer. 2004. “It’s Not Whether You Win or Lose, but How You Play the Game: Self-Interest, Social Justice, and Mass Attitudes toward Market Transition.” *American Political Science Review* 98(3): 437-452.
- Gaines, Brian J., James H. Kuklinski, and Paul J. Quirk. 2007. “The Logic of the Survey Experiment Reexamined.” *Political Analysis* 15(1): 1-20.
- Gartner, Scott. 2008. “The Multiple Effects of Casualties on Public Support for War: An Experimental Approach.” *American Political Science Review* 102(1): 95-106.
- Gerber, Alan S., and Donald P. Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” *American Political Science Review* 94(3): 653-663.
- Gibson, James L. 2008. “Challenges to the Impartiality of State Supreme Courts: Legitimacy Theory and ‘New-Style’ Judicial Campaigns.” *American Political Science Review* 102(1): 59-75.
- Gibson, James L., and Gregory A. Caldeira. 2009. “Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court.” *Journal of Politics* 71(2): 429-441.
- Gibson, James L., and Amanda Gouws. 2003. *Overcoming Intolerance in South Africa: Experiments in Democratic Persuasion*. Cambridge: Cambridge University Press.
- Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. 2000. “Inside the Judicial Mind.” *Cornell Law Review* 86: 777-830.
- Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. 2007. “Blinking on the Bench: How Judges Decide Cases.” *Cornell Law Review* 93(1): 1-44.

(Continued on page 27)

- Harbridge, Laurel, and Neil Malhotra. 2011. "Electoral Incentives and Partisan Conflict in Congress: Evidence from Survey Experiments." *American Journal of Political Science* 55(3): 494-510.
- Huber, Gregory A., Seth J. Hill, and Gabriel S. Lenz. 2012. "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents." *American Political Science Review* 106(4): 720-741.
- Hurwitz, Jon, and Mark Peffley. 2005. "Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Crime Policy." *Public Opinion Quarterly* 69(1): 99-112.
- Jacobson, Gary C. 2012. "The Electoral Origins of Polarized Politics: Evidence from the 2010 Cooperative Congressional Election Study." *American Behavioral Scientist* 56(12): 1612-1630.
- Keeter, Scott. 2011. "Public Opinion Polling and Its Problems." In *Political Polling in the Digital Age: The Challenge of Measuring and Understanding Public Opinion*. Ed. Kirby Goidel. Baton Rouge: Louisiana State University Press.
- Kinder, Donald R., and Thomas R. Pfaffrey. 1993. *Experimental Foundations of Political Science*. Ann Arbor: University of Michigan Press.
- Malhotra, Neil, Joanne M. Miller, and Justin Wedeking. Forthcoming. "The Relationship between Nonresponse Strategies and Measurement Error: Comparing Online Panel Surveys to Traditional Surveys." In *Online Panel Research: A Data Quality Perspective*. Eds. Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Goritz, Jon A. Kroshnick, and Paul J. Lavrakas. Wiley (Survey Methodology Series).
- McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5: 31-61.
- Mintz, Alex, Steven B. Redd, and Arnold Vedlitz. 2006. "Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations?" *Journal of Conflict Resolution* 50(5): 757-776.
- Nyhan, Brendan, and Jason Reifler. 2011. "Blank Slates or Closed Minds? The Role of Information Deficits and Identity Threat in the Prevalence of Misperceptions." Unpublished manuscript.
- Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5): 411-419.
- Peffley, Mark, and Jon Hurwitz. 2010. *Justice in America: The Separate Realities of Blacks and Whites*. Cambridge University Press.
- Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51(3): 515-530.
- Tomz, Michael R., and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107(4): 849-865.
- Transue, John E., Daniel J. Lee, and John H. Aldrich. 2009. "Treatment Spillover Effects across Survey Experiments." *Political Analysis* 17(2): 143-161.
- Vavreck, Lynn, and Douglas Rivers. 2008. "The 2006 Cooperative Congressional Election Study." *Journal of Elections, Public Opinion, and Parties*. 18(4): 355-366.
- Wallace, Geoffrey P.R. 2013. "International Law and Public Attitudes Toward Torture: An Experimental Study." *International Organization* 67(1): 105-140.

Tech Notes

Each edition we will highlight an interesting and useful web page, "app" software, or other form of technology that may be helpful in section members' teaching or research. We welcome reviews and suggestions from section members about programs used in the past that were particularly beneficial and may be of interest to section members.

TESS: Time-Sharing Experiments for the Social Sciences. As mentioned by some contributors in the symposium, this service funded by the National Science Foundation conducts general population experiments for social science researchers. Proposals are screened through a peer-reviewed process and the service is provided for free to those accepted proposals.

<http://tessexperiments.org/index.html>

iCivics: This nonprofit organization was founded by Justice Sandra Day O'Connor and offers lesson plans, interactive games, and other support for civic education classes. While it is somewhat geared toward younger students, the interactive games are informative and can quickly relate important concepts, particularly for introductory-level courses.

<http://frontend-2.icivics.org/>

BOOKS TO WATCH FOR

Drew Lanier, Editor
University of Central Florida
drew.lanier@ucf.edu

Karen J. Alter (Northwestern University and University of Copenhagen) has published *The New Terrain of International Law: Courts, Politics, Rights* (Princeton University Press, 978-0-691-15475-6), which "charts the developments and trends in the creation and role of international courts, and explains how the delegation of authority to international judicial institutions influences global and domestic politics. It presents an in-depth look at the scope and powers of international courts operating around the world. Focusing on dispute resolution, enforcement, administrative review, and constitutional review, Alter argues that international courts alter politics by providing legal, symbolic, and leverage resources that shift the political balance in favor of domestic and international actors who prefer policies more consistent with international law objectives. International courts name violations of the law and perhaps specify remedies. Alter explains how this limited power—the power to speak the law—translates into political influence, and she considers eighteen case studies, showing how international courts change state behavior. The case studies, spanning issue areas and regions of the world, collectively elucidate the political factors that often intervene to limit whether international courts are invoked and whether international judges dare to demand significant changes in state practices."

John E. Finn (Wesleyan University) has published *Peopling the Constitution* (University Press of Kansas, 978-0-700-61962-7). "When most people think of the Constitution, they think of it as a legal instrument, the province of judges and lawyers, who alone possess the expertise and knowledge necessary to discern its elusive and complex meaning. This book outlines a very different view of the Constitution as a moral and philosophical statement

about who we are as a nation. This "Civic Constitution" constitutes us as a civic body politic, transforming "the people" into a singular political entity. Juxtaposing this view with the legal model, the "Juridic Constitution," Finn offers a comprehensive account of the Civic Constitution as a public affirmation of the shared principles of national self-identity, and as a particular vision of political community in which we the people play a significant and ongoing role in achieving a constitutional way of life. The Civic Constitution is the constitution of dialogical engagement, of contested meanings, of political principles, of education, of conversation. *Peopling the Constitution* seeks nothing less than a new interpretation of the American constitutional project in an effort to revive a robust understanding of citizenship.."

Cesare Romano (Loyola University), **Karen J. Alter** (Northwestern University and University of Copenhagen), and **Yuval Shany** (Hebrew University), (eds.) have published *The Oxford Handbook on International Adjudication* (Oxford University Press, 978-0-19-966068-1). "The post-Cold War proliferation of international adjudicatory bodies and international adjudication has had dramatic effects on both international law and politics, greatly affecting international relations, particularly economic relations, the enforcement of human rights, and the criminal pursuit of perpetrators of mass atrocities. International courts and tribunals have become, in some respects, the lynchpin of the modern international legal system. *The Oxford Handbook of International Adjudication* uniquely brings together analysis of the legal, philosophical, ethical and political considerations brought about by these bodies. It provides an original and comprehensive un-

(Continued on page 29)

derstanding of the various forms of international adjudication. A series of cross-cutting chapters overview key issues in the field, both theoretical and practical, providing scholars, students, and practitioners with a detailed understanding of important legal and political influences within the international adjudicative process.”

David Schultz (Hamline University) has published *Election Law and Democratic Theory* (Ashgate, 978-0-7546-7543-3). “This book provides a full-length examination of the political theories

and principles and democratic values underlying current election law debates and the regulation of political campaigns and participants in the United States. Topics covered range from campaign finance reform, voting rights, reapportionment and ballot access to the rights of political parties, the media and other players in the system. It challenges much of the current debate in election law and argues for more discussion and development of a democratic political theory to support and guide election law jurisprudence.”

Rachel VanSickle-Ward (Pitzer College) has published *The Devil is in the Details* (SUNY Press, 978-1-4384-4923-4). “The level of detail in a given law can have dramatic consequences for how that law is interpreted and applied. In *The Devil Is in the Details*, VanSickle-Ward focuses on the dynamics of social policy construction in the United States in order to better understand why the wording of legislation can range from the specific to the ambiguous. When policies are [highly salient,] the fissures produced by partisan discord, interest group diversity, and pluralistic executive branches promote ambiguous policy. When policies are lower profile, this relationship is more tenuous and, at times, inverted, with contention producing more policy detail. Put simply, on important and controversial legislation, ambiguity serves as a vehicle for compromise when key participants disagree over details. Moreover, fragmentation is a more powerful driver of ambiguity than limits in technical expertise or legislative capacity. This multi-method investigation takes a novel approach to the measurement of statute specificity. The author combines comprehensive content analysis of more than 250 health and welfare bills passed in 44 states in the 1990s and 2000s with in-depth interviews of policy-making elites.”

ANNOUNCEMENTS

Call for Papers & Upcoming Conferences

- Midwest Political Science Conference (April 3-6)
- New England Political Science Association (April 24-26)
- Law and Society Association: (May 29-June 1)
- Northeast Political Science Association: (November 13-15)
 - Proposal Deadline: June 14, 2014