

**CONFERENCE ON LABORATORY EXPERIMENTS TO ESTABLISH A CONSORTIUM FOR
LABORATORY EXPERIMENTS IN POLITICAL SCIENCE**

Cindy D. Kam, Vanderbilt University

Version: 8/23/10

The experimental approach has begun to permeate political science research, especially in the last decade. Research universities across the country have established social science laboratories for experimental research, graduate students receive training in conducting and implementing their own experiments, and Time-Sharing Experiments in the Social Sciences (TESS) was developed to provide a peer-reviewed process by which experimentalists could place small modules on surveys administered to national samples of respondents. An entire issue in the 2002 volume of *Political Analysis* was devoted to experimental research methods in political science, and a recent article in the *American Political Science Review* was devoted entirely to “The Growth and Development of Experimental Research in Political Science” (Druckman et al. 2006). Later this year, an entire handbook on experiments in political science (*The Handbook of Experimental Political Science*) will be published by Cambridge University Press (Druckman et al. 2010).

Appropriate experimental designs offer researchers unparalleled control over stimuli, wide latitude in measurement, and, above all, *internal validity*: the ability to identify causal relationships between a stimulus and a response.¹ Conducting experiments in a laboratory setting provides maximum assurance of internal validity, due to the superior control over environmental conditions laboratories offer. Moreover, laboratory settings may be preferred due to the wider range of questions, tests, and tasks that can be administered compared with studies in the field. For example, researchers who are interested in collecting millisecond response latency data, who want to administer sophisticated audio-visual treatments, or who wish to employ psychophysiological measures, may prefer or need to conduct studies in the controlled laboratory setting rather than the field. Laboratory experiments also offer researchers greater *control* over treatment administration and study implementation, as subjects are often under the direct supervision of trained research assistants, are unlikely to be doing other things while answering questions, and are more likely to comply with instructions. Furthermore, researchers may also rely on laboratory-based studies because other modes of data collection (e.g., fielding a nationally-representative survey) are significantly more costly. However, laboratory experiments are often criticized for their lack of generalizability and can suffer from a power limitations if the researcher’s local subject pool is limited in size.

This document proposes a Conference on Laboratory Experiments organized with two key purposes in mind. The primary purpose is to establish a Consortium for Laboratory Experiments for Political Science. Such a Consortium would (a) create a subject trading pool whereby principal investigators from various laboratories could trade out modules to be administered at different locales with different subject pools; (b) facilitate research collaborations across experimental laboratories; and (c) provide faculty and graduate students who are not at institutions with active experimental labs to purchase or compete for access to a network of experimental labs. The establishment of this Consortium would enable laboratory researchers to enhance the *external validity*

¹ By “stimulus” we refer to the experimental manipulation(s) typically introduced by random assignment to experimental subjects. We expect a response to that stimulus, and thus, the stimulus is typically seen as the “cause” and the response seen as the “effect.”

of their studies by testing for the generalizability of their causal claims across distinct individuals, settings, contexts, and operationalizations. The Consortium would also enhance the ability of researchers to test more nuanced and complex second-generation theoretical questions about the extent to which, and the conditions under which, some key causal variable affects important political outcomes. Finally, the Consortium would broaden access to laboratory experiments, enabling both faculty and graduate student researchers to have access to experimental laboratories. The secondary purpose is to disseminate research articles featuring innovative techniques employed in laboratory experiments.

THE GROWTH OF EXPERIMENTATION IN POLITICAL SCIENCE

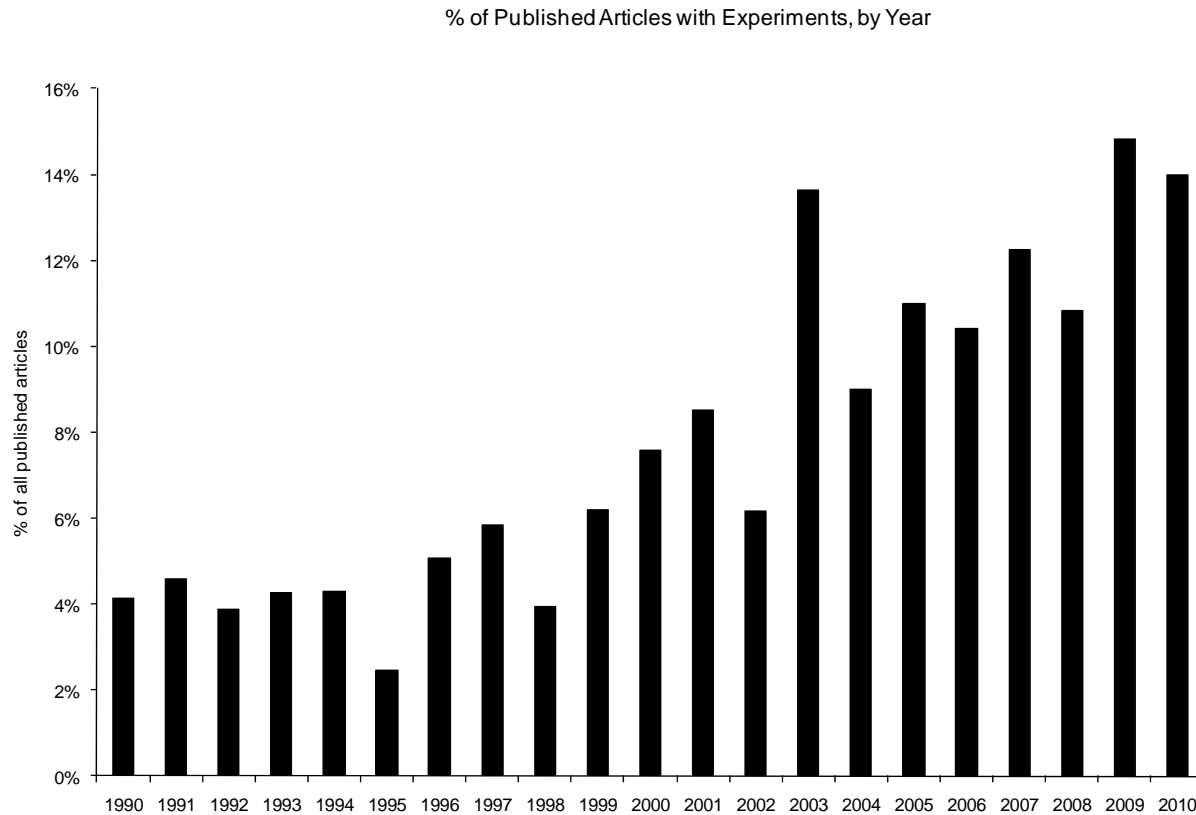
By stepping outside a world in which “causal connections subtly intertwine” (Achen 1992, p. 196) into a controlled situation, experimental researchers gain an unparalleled ability to isolate cause-effect relationships. Achieving such internal validity is “the *sine qua non*” (Campbell and Stanley 1963, p. 5) and, at times, the end goal of experimental research (see, e.g., McDermott 2002a for a forceful discussion on this point). Experiments offer more than an unrivaled ability to disentangle cause from effect. Experiments can adjudicate in disputes raised by observational data, they can test underlying assumptions in theories, and they can inform policymakers (Druckman et al. 2006, p. 629). Further, by isolating cause and effect, experiments can be very useful in theory testing, and indeed, for many researchers, theory testing is the primary purpose of experimentation (Aronson, Wilson, and Brewer 1998: 133).

Political scientists have increasingly opted to incorporate the experimental method into their work. Figure 1 depicts a content analysis of articles published in five political science journals: the three top-tier general-interest journals (*American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*) and two specialized journals (*Political Behavior* and *Political Psychology*), from 1990–August 2010. In the twenty-year period in Figure 1, we see a trajectory of growth in publication of experimental studies.² Experiments are growing in prominence across both general-interest and specialized journals in political science.³

² I follow Aronson, Wilson, and Brewer (1998) and consider an “experiment” to be a study in which subjects are randomly assigned to receive some manipulation. Experiments conducted in either the field or the laboratory are eligible for inclusion under this category. “Natural experiments” and “quasi-experiments” are not included (see King, Keohane, and Verba 1994, p. 7, note 1, for a critique of the latter term). Observational studies do not count as experiments (i.e., standard public opinion surveys without embedded experiments, personality assessments, participant observation, etc.).

³ This finding of an increased number of articles featuring experiments is consistent with other content analyses. McDermott (2002b) reports an upward trend based on her review of a number of political science journals from 1926 to 2000. Morton and Williams (2010) restrict their review to the *American Political Science Review* (APSR), *American Journal of Political Science*, and *Journal of Politics* between the years 1950 and 2007; they find that publication of experiments “has increased at an astonishing rate” (p. 3-4). Finally, Druckman et al. (2006) document a similar positive trend in their analysis of the APSR and its contents from 1906 to 2004. Moreover, the authors examine the citation rate (via the Web of Science Social Sciences Citation Index) for the experimental articles they find in comparison to a sample of nonexperimental articles. They find that the studies with experiments are cited more frequently. Thus, not only are experiments on the rise but their impact is presumably increasing in step with that rise.

Figure 1. Prevalence of Published Experiments in Political Science Journals, 1990-2010



Note: Journals include: *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Political Behavior*, & *Political Psychology*, 1990 - August 2010.

Political scientists rely on a range of types of experiments: laboratory experiments, survey experiments, and field experiments, in particular. While there are various advantages to survey experiments and field experiments, neither type of experiment offers as strong a case for internal validity as laboratory experiments, where internal validity focuses on the causal relationship between the experimental treatment and some outcome. Indeed, when it comes to the ability to offer strong causal evidence, McDermott notes, “laboratory experiments are the gold standard” against which other types of experiments are evaluated (2002, p. 32). Still, experiments, and laboratory experiments in particular, find themselves vulnerable to criticisms of external validity, where external validity concerns the ability to generalize the cause-effect relationship. As Campbell and Stanley (1963) put it, external validity “asks the question of *generalizability*: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?” (p. 5)

EXTERNAL VALIDITY: THE ACHILLES’ HEEL TO LABORATORY EXPERIMENTS

Skepticism of the experimental method is primarily lodged in concerns regarding external validity (Lupia 2002: 320). Laboratory experiments based upon student samples frequently face this criticism, and political scientists who conduct laboratory experiments rely substantially on student subjects. As Kam, Wilking, and Zechmeister (2007) report in their content analysis of major political science journals from 1990-2006, in the top-tier, general-interest journals, about a quarter of experiments are based solely on student samples; in the specialized journals (*Political Behavior* and

Political Psychology), 70% of the experiments utilize student subject pools. Use of student samples is widespread among laboratory experimentalists, and it is, by some accounts, an Achilles' heel.

The point of particular concern involves generalization from the sample of experimental participants—especially when, as is often the case, the sample consists of students—to a larger population of interest. Indeed, this was the focus of Sears' (1986) widely cited article, "College Sophomores in the Laboratory: Influences of a Narrow Data base on Social Psychology's View of Human Nature." Many political scientists employ "the simplistic heuristic of 'a student sample lacks external generalizability'" (Kam, Wilking, and Zechmeister 2007, p. 421). Gerber and Green (2008) similarly note that "If one seeks to understand how the general public responds to social cues or political communication, the external validity of lab studies of undergraduates has inspired skepticism (Sears 1986, Benz and Meier 2006)" (p. 358). In short, social scientists in general and political scientists in particular view student subjects as a major hindrance to drawing inferences from experimental studies.

Yet, student samples are *not necessarily* inferior to other samples (see Druckman and Kam 2010 for an extensive demonstration of this point).⁴ The appropriateness of the sample depends on what is being studied, and researchers have the obligation to think through the advantages and disadvantages of not only the type of experiment employed but also the subject pool recruited. First, student subjects are appropriate in cases where there is theoretical and/or empirical reason to believe that the effect of the treatment would be the same (or closely similar) across student and nonstudent samples. Second, student samples might be appropriate in cases where student samples provide a more critical test of the hypothesis. Typically in such cases scholars argue that certain characteristics of their student sample make the use of that sample a more stringent test of their hypothesis(es) (see, for example, Sigelman, Sigelman, and Bullock 1991, p. 135; Kahn and Geer 1994, p. 100; Funk 1997, pp. 690-691). The expectation is that the effect of the treatment will be lowest among this sub-group; thus, if results are obtained one can assume they would obtain, and likely at a greater level, in a more general population. Alternatively, at least one set of scholars (Druckman and Nelson 2003) argues that the effect of their treatment is likely to be greater among their student sample and, because the research is testing the hypothesis that there will be no effect, the student sample provides a more difficult test. In these cases, the relative convenience, ease, and cost of recruiting undergraduate subjects make this subject pool not only an attractive but an efficient solution. Third, student subjects are appropriate when students are the population of interest. For example, Ishiyama and Hartlaub (2002) carried out a study with student subjects, with the intention of determining whether or not students were sensitive to differences in the ways in which political science syllabi are worded. And, fourth, student subjects are appropriate when the principal purpose of the study is to investigate a specific, theoretically interesting causal relationship; in other words, when the focus is on internal validity. For example, in experimental research using student subjects to study distributive justice, Scott et al. (2001, p. 762) argue, "Whatever the limits of external validity, however, our first concern was internal validity. Our primary aim was to examine the relationship among allocation principles and to see how they changed when we manipulated theoretically relevant conditions."

Furthermore, assessment of any single study, regardless of the nature of its participants, must be done in light of the larger research agenda to which it hopes to contribute. And, when it comes to

⁴ For a discussion of when student subjects might not be appropriate, see Kam, Wilking, and Zechmeister (2007).

generalization from a series of studies, the goal is to generalize across *multiple* dimensions. External validity refers to generalization not only across types of individuals but also across settings/contexts, times, and operationalizations. External validity is best understood as being assessed over a *range of studies* on a single topic (McDermott 2002, p. 335). Liyanarachchi (2007, p. 55) explains:

According to experts on methodology, true external validity of findings can only be obtained by converging the results of many studies in an area... Reiterating this point in social sciences, McGrath et al. (1982: 105) suggested: “No one ‘finding’ is evidence, and no one study yield[s] “knowledge;” empirical information can gain credence *only* by accumulation of convergent results.

Laboratory researchers may not want to tradeoff the internal validity of laboratory experiments for external validity: as Anderson and Bushman (1997) note, “if a study has low internal validity—if it doesn't clearly demonstrate a causal relation between the independent and dependent variables—then there is nothing to generalize” (p. 21). Still, external validity looms large as a concern in conducting, evaluating, and publishing experimental research, particularly for laboratory experimentalists.

The establishment of a Consortium for Laboratory Experiments in Political Science would enable laboratory experimentalists to address external validity concerns. Such a Consortium would (a) create a subject trading pool whereby principal investigators from various laboratories could trade out modules to be administered at different locales with different subject pools; (b) facilitate research collaborations across experimental laboratories; and (c) provide faculty and graduate students who are not at institutions with active experimental labs to purchase or compete for access to a network of experimental labs.

A CONFERENCE TO ESTABLISH A CONSORTIUM FOR LABORATORY EXPERIMENTS IN POLITICAL SCIENCE

The Conference on Laboratory Experiments in Political Science is tentatively scheduled for May 5-6, 2011 at Vanderbilt University. The purpose of the Conference on Laboratory Experiments would be two-fold. Its primary purpose would be to establish a **Consortium for Laboratory Experiments in Political Science**. Its secondary purpose would be to foster the dissemination of innovative experimental research conducted in political science research laboratories.

The two-day conference will feature morning panel sessions devoted to presentation of innovative research conducted in experimental research laboratories. The afternoon sessions will be devoted to focused and extended discussion regarding the Consortium for Laboratory Experiments in Political Science. I expect attendees to arrive the evening of Wednesday, May 4 and depart the evening of Friday, May 6 or the morning of Saturday, May 7.

Purpose I: Establish a Consortium for Laboratory Experiments in Political Science

Rationale for a Consortium: As detailed above, laboratory studies face challenges when it comes to external validity. Even though laboratory experimentalists are primarily interested in identifying causal mechanisms, many laboratory experimentalists (and many of their reviewers) also seek to “generalize” a documented causal relationship. The Consortium would enable a researcher to test for generalizability across a range of studies, in a number of ways.

At the most basic level, “generalizing” a relationship could mean examining the extent to which a given experiment (with identical stimuli, procedures, etc.) yields the same results when run on a different group of subjects. As a practical matter, many experimental researchers run into a problem implementing this advice, given a limited number of subjects in the experimental pool. Re-running a study at the same institution might mean that the same subjects who initially participated in the first study might participate in the second as well, thus hampering the researcher’s ability to conduct a truly independent test. Moreover, depending upon the extent to which subjects were debriefed about the purpose of the first study (often in accordance with IRB requirements and also with the expressed purpose of educating subjects about the experimental method and about the study in particular), they might be more likely to play along with the researcher’s hypotheses (Orne 1962) or they might sabotage it if they were to take it again. In either case, the conditions are not ideal for a second, independent investigation. Establishment of the Consortium would allow for such tests, since laboratory researchers could trade subjects and run the same study at a different university, with a different subject pool, at a different point in time, to determine the generalizability of the results.

However, external validity does *not* simply refer to whether a specific study, if re-run on a different sample, would provide the same results. It also refers more generally to whether *conceptually equivalent* relationships can be detected across people, places, times, and operationalizations (Anderson and Bushman 1997). Hence, within the Consortium, a researcher could probe the extent to which a causal relationship is generalizable under different operationalizations of the key independent or dependent variables. In this case, a researcher could run the second study or third study at the home institution, but the researcher might prefer to run the second study at a different university, in order to tap into a fresh and truly independent subject pool that has not been previously exposed (or debriefed about) a particular study’s purposes.

Aside from these external validity concerns, replication and extension of experimental studies offer the potential for theoretical and empirical progress. The Consortium would allow researchers to gain access to a different set of subjects: a set of subjects that are, on average, different in theoretically important and potentially revelatory ways. This would enable researchers to develop more complex and nuanced theoretical expectations about the ways in which a particular treatment effect might be contingent upon particular individual characteristics: enabling a researcher to move from a first-generation question of “Does X cause Y?” to a second-generation question of, “Under what conditions, for which individuals, within which contexts, does X cause Y?” (Kam and Franzese 2007). For example, a researcher at a predominantly conservative institution might want to tap into a subject pool at a predominantly liberal institution, to ascertain whether treatment effects are contingent upon ideological predispositions. A researcher studying race or ethnicity might want to tap into a subject pool at more racially diverse institutions, pooling subjects across institutions, thereby allowing for the systematic study of under-represented groups. A researcher interested in context effects could examine the extent to which contextual features such as urban/rural, public/private, secular/nonsecular characteristics interact with particular treatments. To the extent that interesting interactions between treatment effects and characteristics appear, this would provide theoretical and empirical nuances to what had been straight-forward question about whether X causes Y. To the extent that interesting interactions do not appear, then this would provide support for the generalizability of the causal relationship that a researcher has uncovered.

Finally, the Consortium would enable researchers with limited subject pools to overcome some statistical power constraints by enlarging their sample size. This, too, would enable researchers to

explore the nuances of their theories, by conditioning treatment upon theoretically important covariates. Such interactive analyses may be difficult or indeterminate given limited sample size (and limits on the sample range of any key covariate). Pooling subjects across sites would enable researchers to examine conditional treatment effects with greater statistical leverage.

Conference Attendees: The PI plans to identify a diverse group of 20 researchers from institutions in the United States and abroad who are affiliated with political science experimental laboratories. These researchers would be invited with an eye towards diversity in subfield, research expertise, rank, gender, race/ethnicity, and characteristics of their institutions. Moreover, the PI plans to invite international scholars to broaden the reach of the Consortium beyond the United States and to encourage cross-national collaborations.

Background Preparation: Compilation of Information on Research Labs: Prior to the conference, the PI and graduate student assistant would create an online survey that inquires into the details of each of the experimental laboratories of the attendees: size, capacity, subject pool, subject characteristics, availability for experimental use, staffing, software, protocols, IRB requirements, and types of projects that have been fielded in the laboratories. The PI and graduate student would compile this information into a report and disseminate it ahead of time to conference attendees, so all would have common knowledge about the different labs that would be represented at the conference.

Tasks for the Consortium: Significant portions of the conference would be devoted to developing concrete rules and procedures for the Consortium. Plenary sessions would be devoted to distinct topics.

- *Mission:* The attendees would craft the Consortium's mission statement.
- *Membership and Fees:* The attendees would discuss membership requirements, including fees, and tangible benefits that would be given to laboratories that were members of the Consortium.
- *Trades:* We would discuss the logistics of how to trade subjects and research modules across research labs. In particular, we would identify an algorithm for quantifying "trades" across laboratories in a fair and equitable manner. This would likely be based upon length of proposed module (number of questions, or time required) as well as sample size and any special procedures that would need to be administered.
- *Logistics of Study Implementation:* We would discuss issues regarding subject recruitment and subject pools; software; programming; and IRB requirements.
- *Common Practices:* We would delineate common practices to ensure that research studies were implemented as consistently as possible across sites.
- *Data Security, Sharing and Storage:* We would identify the manner in which data would be shared, stored, and secured across laboratories.
- *Action Plan:* We would develop an action plan to initiate the first cross-university subject trades, to be implemented in the Fall 2011 semester.
- *Broadening Access to Researchers at Institutions that Do Not Have Laboratories:* We would also develop procedures through which researchers (faculty or graduate students) at institutions that do not house experimental labs could gain access to the Consortium. For faculty-initiated research projects, this would likely be in the form of either explicit collaboration or a purchase agreement. For graduate students, this could take the form of an open Request for Proposals (to be disseminated widely via the organized sections from the national and

regional political science associations), through which graduate students could submit proposed studies that would be reviewed by a panel of Consortium members. The fee structure for faculty-initiated research projects might be designed such that it could subsidize the costs of graduate student research projects; any remaining costs might be subsidized by the hosting laboratory or Consortium membership fees, or supplemented by research funds provided by the student.

Dissemination of Information: Following the conference, the PI and graduate assistant would disseminate information about the Consortium via the internet. A web presence would be created, which would allow for online submission form to propose trades across member laboratories, as well as online submission form for faculty and graduate students who are not in member institutions but would like access to experimental lab facilities.

Purpose II: Disseminate Research Featuring Innovations in Laboratory Experimentation

A secondary purpose of the conference would be to bring together a diverse range of scholars engaged in laboratory experiments, to showcase innovations in measurement and implementation. Social scientists have increasingly adopted innovative techniques, including psychophysiological measures such as eye-blink startle reflexes (Oxley et al. 2008), eye-tracking (e.g., Galesic et al. 2008), and skin conductance responses (Mutz and Reeves 2005; Oxley et al. 2008), as well as biochemical measures such as testosterone levels (McDermott et al. 2007). The use of these measures is nearly entirely limited to laboratory studies. Political scientists have also increasingly adopted the use of implicit measures that rely upon subliminal priming (Kam 2007a; Kam and Zechmeister 2010; Burdein, Lodge, and Taber 2006; see Lodge and Taber 2010 for a review), a technique that is not necessarily limited to laboratory studies but is commonly employed only in such locations. Additionally, studies that use decision-boards to study information-seeking (Lau 1995; Lau and Redlawsk 1997, 2006; Redlawsk 2004) rely primarily on the ability to have subjects use a personal computer that has the necessary software loaded onto it (which either means having subjects come to a laboratory or bringing computers to subjects). During these panel sessions, the conference would be to bring together a group of laboratory experimentalists to share these innovations in measurement and to discuss the logistics of employing such measures.