ANUARIO CIEP
Universidad de Costa Rica, ISSN: 2215-2873

EXPERIMENTOS EN LA CIENCIA POLÍTICA: APORTES Y DESAFÍOS A LA DISCIPLINA

EXPERIMENTS IN POLITICAL SCIENCE: CONTRIBUTIONS AND CHALLENGES TO THE DISCIPLINE

Ronald Alfaro-Redondo

Correo electrónico: ralfaro@estadonacion.or.cr

Recibido 17/03/2014 – Aceptado 03/02/2015

Political Scientist from Costa Rica. PhD Candidate. Political Science Department. University of Pittsburgh. Pennsylvania, USA. Master of Arts in Quantitative Methods in the Social Sciences. Columbia University, New York. USA. Master of Arts in Political Science. University of Pittsburgh. Pennsylvania, USA. Principal researcher Americas Barometer Costa Rican chapter. Researcher in Programa Estado de la Nacion. Professor in the Political Science Department. University of Costa Rica.

RESUMEN

En este "review article" se describen las ventajas de los experimentos en la ciencia política actual y se resume el estado de la cuestión en la literatura especializada, la cual apunta que el aumento que presenciamos se debe, fundamentalmente, al hecho de que los experimentos aventajan a los otros métodos en dilucidar las relaciones causa-efecto. Los experimentos han demostrado ser lo suficientemente flexibles para adaptarse a las preferencias de los investigadores, las preguntas complejas y la innovación en los diseños. Por ello, hemos presenciado la expansión de los métodos, los avances en la depuración de teorías, y el progreso en el establecimiento de relaciones causales que los experimentos han traído a la ciencia política. En el último apartado se mencionan las principales limitaciones de los diseños de investigación experimentales.

Palabras claves: experimentos, diseños de investigación experimentales, validez, causalidad, generalización.

ABSTRACT

In this review article I describe the advantages of experimental research in political science and summarize the state-of-the-art in the literature that postulates that their proliferation is mainly because experiments are superior to all other methods in disentangling cause and effect relationships. Experiments have proven to be versatile enough to accommodate into researchers' preferences, complex research questions and innovative designs. We have witnessed the expansion of methods, the advance in theory testing, and the improvement in establishing causal relationships that experiments have brought to political science. Also, I provide some examples of experiments in the literature that expanded our knowledge. In the last section I offer an overview of the limitations of experimental design.

Keywords: experiments, experimental design, validity, causality, generalizability.

INTRODUCTION

In the mid-1920s an acute researcher, Harold Gosnell, inspired mainly by intuition, was exploring new avenues for understanding voting behavior. In the context of presidential elections Gosnell conducted an experiment. He was concerned with a simple issue: would individuals who get information ("treatment condition") regarding the election (e.g. registration deadlines and polling stations locations) tend to vote at a higher rate than others that did not receive it ("control group")? Ten years later (1935) George Hartmann conducted another experiment analyzing the relationship between rational or emotional appeals and supporting the Socialist Party. Even though their concerns were conventional, their methods were not. In that sense, Harold Gosnell and George Hartmann delineate the foundations of what later became one of the most promissory fields in the study of political behavior (Kinder and Palfrey, 1993).

In this essay I summarize some of the reasons why we have been seen an increasing trend in the number and importance of experiments in the discipline. Also I discuss the good things experimental research has brought to political science, things that we would not have learned otherwise. Also, I provide some examples of experiments in the literature that expand our knowledge. At the end I pay attention to some limitations of experiments.

EXPERIMENTS "ARE IN THE AIR"

In a discipline highly dominated by surveys, archival, and other observational studies, Gosnell and Hartmann's contributions inspired others for studying mass behavior through a more plural and diverse methodological approach. Nowadays experiments are increasingly being used for understanding a broad array of topics: voting behavior, political attitudes, stereotypes, and media effects. Different subfields have implemented experimental studies in their own specialization areas. This trend has been more visible in the last two decades. More experimentally-based papers published, calls for specialized conferences, several chapters in both the Handbooks of *American Elections and Political Behavior* and *Political Methodology*, as well as the recent publication of the *Handbook of Experimental Political Science* and the *Journal of Experimental Political Science*, all these exemplify the increasing importance of experimental research. Also, as the diversity of experiments demonstrates, they have proved to be versatile enough to accommodate into researchers' preferences and likes, research questions and designs. Today, someone interested in conducting experiments has available an interesting range of experiment types for choosing depending on the phenomenon they want to study. This trend has, without doubt, been beneficial for both the discipline and its political scientists. Since Gosnell path-breaking study, we have witnessed the accumulation of knowledge, the expansion of methods, the

advance in theory testing, and the improvement in establishing causal relationships that experiments have brought to political science.

Fridkin and Kenney (2012) identified several reasons why research in political behavior, especially in public opinion, has moved in an increasingly experimental direction. First, experiments, unlike other methods, can provide authoritative answers to causal questions. Second, the random assignment of subjects to stimulus and control conditions holds constant all potentially confounding forces. Third, because of the control inherent in experimentation, experiments are superior to all other methods at sorting out and disentangling cause and effect relationships. Fourth, with experiments, researchers can untangle complex phenomena, to sort out the details of the underlying process. Fifth, experiments are often relatively simple and economical to execute. A single researcher, for example, can design and implement an experiment, or even several of them, with a relatively limited budget. Finally, when experiments are properly designed, the data analysis is clear-cut, allowing researchers to present their findings in a straightforward and easy to comprehend manner.

Although there are a number of reasons for the increase in experimentation in political science, the dominant explanation, according to Morton and Williams (2008:340), for the expansion since 1990

> …is the increase in cheap and easily programmable computer networking technology, for the number of possible experimental designs via laboratory networks and over the internet greater than what researchers can conduct manually.

Also they point that

> …computer technology has also led to a greater ability to engage in survey experiments, and to deal with the statistical and other methodological issues that are involved in field and natural experiments.

In sum, "technology has facilitated the conditions for the transformation of political science into an experimental design".

WHAT ARE EXPERIMENTS AND WHAT GOOD THINGS THEY HAVE BROUGHT US?

Campbell et al. (1963) define experiments as the portion of research in which variables are manipulated and their effects upon other variables observed. For Kinder and Palfrey (1993) experimentation may refer to a single form of scientific inquiry. Experiments are undertaken in the laboratory, in the field and even

rarely under natural circumstances. Experiments test the response of individuals, groups, neighborhoods, organizations, or cities. No matter how diverse experiments may be in practice, they share an interventionist spirit. Experiments

> …*intrude upon nature*, and they do so (almost always) to provide answers to causal questions. It is the feature of intervention, and the control that such intervention brings, that, distinguishes experimental research from other systematic empirical methods (Kinder and Palfrey, 1993, p. 6).

Morton and Williams (2008, p. 341) provide a more elaborated definition. For them, the defining characteristic of experimental research is "…intervention by the researcher in the data-generating process". In experimental research, the variation in the data is partly a consequence of the researchers' decisions at the design stage before data are measured. Consequently, non-experimental empirical research involves using only data in which all the variation is a consequence of factors outside of the control of the researcher. In other words, the researcher only observes the data-generating process, but does not intervene in that process.

In contrast with other types of research, experiments often require small numbers of participants and may even rely only on student participants. As a result, sometimes participants are not representative of the entire population, especially those experiments using undergraduate students (Sears, 1986). The choice of participants may not matter for some topics, but using college students limits the results to people who are used to dealing with abstract concepts and more complex phenomena (Niemi, 2011). Because of this, questions like what can we believe about what we learned or how valid is the research? are common in the experimental design field. These questions lead us to the discussion of the *validity* of experimental political science research.

Traditionally experiments have been criticized in terms of what is known as *external validity,* that is, whether we can generalize experimental findings. External validity "…means whether causal inferences established in an empirical analysis hold over variations in persons, settings, treatment variables, and measurement variables"(Morton and Williams, 2010, p. 344). Concretely: Do the results from one data-set generalize to another?

Experiments have contributed to the theoretical and methodological debate regarding research validity. Without them we would not have improved our knowledge in this crucial aspect. Otherwise, misunderstandings and misinterpretation would have prevailed. For instance, some scholars argue that it is a mistake to equate external validity with whether a given data-set -used to establish a particular causal relationship- resembles the unmanipulated data-generating process. Instead, establishing whether a result is externally valid involves replication of the results across a variety of data-sets. In sum, external validity is

> …really about the robustness of the experiments across different formulations, and not about whether the experiment resembles the hypothesized unmanipulated data-generating process" (Morton and Williams, 2010, p. 345).

Another important thing we have learned from experiments is that they are better able to establish causation, as the researcher manipulates the experimental variable and control (or randomizes) for the effects of extraneous variables (Niemi et. al. 2011). The experimental goal is, essentially, to understand causality, causal relationships, and provide a much stronger grasp about causal effects. Therefore, experiments are considered the gold standard in terms of causality. Also, experimental designs allow us to address one of the necessary condition for defining causal relationships: spuriousness[1].

The capacity of experiments to provide decisive tests of causal propositions follows from two aspects of control emphasized in experimental practice. First, by creating the treatments of interest, the experimenter holds extraneous factors constant and ensures that subjects encounter treatments that differ only in designated ways. Second, by assigning subjects to treatments randomly, the experimenter can be confident (within the limitations established by statistical inference) that any differences observed between subjects assigned to different treatment conditions must be caused by differences in the treatments themselves (Kinder and Palfrey, 1993).

In addition, the development of experimental research has demystified the strengths of the surveys, one of the most common methods in the discipline, to address causal questions. A causal question involves, according to Druckman et al. (2011, p. 16), a comparison between two states of the world: one in which some sort of intervention is administered (a treated state, i.e. exposing a subject to a stimulus) and another one in which it is not (untreated state). The fundamental problem of causal inference arises, they argue, because we cannot simultaneously observe an individual in both its treated and untreated states. For solving this issue, they continue,

> …surveys employ an underlying logic: identify a group of comparable observations (individuals with the same or similar age or ideology) that have received different treatments then conduct the causal evaluation primarily or exclusively on these observations. However, the problem is that these approaches fail to eliminate comparability problems. More specifically, two groups of individuals who look the same to researchers could differ in unmeasured ways.

1 The other ones are covariation and temporal ordering.

Experiments possess a powerful attribute for dealing with these issues. Experimental research differs from surveys (and other observational studies) in that the subjects under study are randomly assigned to different treatments (Druckman et al. 2011).

So, because of the belief in the inability of survey data to answer causal questions, some researchers have turned to experimental methods (Morton and Williams, 2010). For example, Lodge et al. (1995) and Lau and Redlawsk (2006) use experiments to evaluate how voters integrate information into evaluations in political contexts. Also, a number of researchers have begun to use experiments to explore the assumptions about individual behavior underlying game theoretic and rational choice models in an effort to better understand political choices. Others use experiments to test whether individuals rationally interpret the use of arguments by other speakers while engaging in deliberation (Dickson, 2008). Again, as these cases illustrates, these are research questions that individual surveys, archival data or other kind of methods, are incapable to answer.

Kinder and Palfrey (1993) emphasize in another important contribution associated with the implementation of experimental political science research, called *analytic decomposition*. By creating treatment and control conditions, the experimenter is able to isolate one causal variable at a time. This, in turn, allows complex phenomena to be decomposed in a way that is impossible under other research strategies. Thus, the basic purpose of an experiment is to isolate the causal influence of different interventions (Green et al. 2012).

GOOD EXAMPLES OF EXPERIMENTS IN THE LITERATURE

In this section of the article I describe the main features and findings of three outstanding experimentally-based researches of three different kinds: survey, lab, and field experiments. The selection of the cases was based on the salience of the findings in terms of both the expansion of our knowledge and the challenge of the conventional view. First I summarize Sullivan et al. (1978), then I describe Gerber et al. piece (2003), and finally I discuss Mutz and Reeves (2005).

One of the major controversies in the field of political behavior concerns the degree of the electorate's ideological awareness. In a work published in the sixties Philip Converse (1964) provides evidence indicating that there was little stability and continuity in most voters' thinking. This image of the uninformed and unsophisticated voter began to reshape the view of the citizenry and democratic politics. Following Converse's arguments, scholars argued that in the 1960s emerged a more coherent and ideologically consistent set of attitudes in the American public (Nie and Andersen, 1974).

Using an experimental design in two independent surveys in which one sample was asked for their views using the pre-1964 question format; and the other sample was asked using the question format introduced in 1964, they found good reasons to believe that the thesis of changes in the structure of public attitudes had been exaggerated. Sullivan et al. (1978) findings show that the level of constraint in the mass public has not increased greatly over time, as others have argued, but rather it merely appears to have increased because of the ways in which it has been measured. Concretely, changes in measurement instruments rather than real changes in the structure of public attitudes are probably the real cause of the pattern of correlations reported by others (Sullivan et. al. 1978).

Sullivan et al. (1978) piece is an example of survey experiments. There are several types of survey experiments. For instance, experiments embedded in survey questionnaires, telephone, and web-based modes. Survey experiments have undergone radical transformations due to the impact of technology and the diversification of the instruments available (computer-assisted interviewing technology).[2] These changes allow scholars to combine multiple characteristics of the subjects and manipulate many more conditions than printed questionnaires.

"Field experiments" refer to experiments that take place in real world settings (Green et. al. 2012). Political science field experiments take many topics including campaign finance, mobilization, and lobbying. The recent revival of field experiments in the discipline began with a series of experimental studies of campaign activity (Gerber and Green, 2000) and voting behavior (Gerber et. al. 2008). Drawing on a field experiment involving 25,200 registered voters conducted prior to the general election of 1998 (Gerber et. al. 2003) provide the first direct test of the hypothesis that casting a ballot in one election increases one's propensity to go to the polls in the future, a behavior denominated *habitual voting*. In this experiment subjects were randomly assigned to treatment conditions in which they were urged to vote through direct mail or face-to-face canvassing. Compared to a control group that received no contact, the treatment groups were significantly more likely to vote in 1998. Consequently, voting in one election substantially increases the likelihood of voting in the future. Indeed, according to the authors, the influence of past voting exceeds the effects of age and education reported in previous studies.

Analyzing habitual voting through experimental designs is relevant because under circumstances of substantial turnout reductions, like in the industrial democracies in the last decades, the development of electoral habits turns to be crucial due to the possibility that many more voters cultivate 'habitual voting' may constitute a key factor to revert such tendency in the medium term or in the long run. Nevertheless, as the number of individuals that do not cast their vote increases, lower turnout would predominate worsening political participation gaps among

2 Piazza, Sniderman and Tetlock (1989) describe some of the changes in their paper.

voters on one hand, and undermining political representation on the other. Experiments can help us to elucidate these questions.

Finally, laboratory experiments are experiments where the subjects are recruited to a common location, the experiment is largely conducted at the location, and the researcher controls almost all aspects in the environment in that location, except for subjects' behavior. Morton and Williams (2008) discuss some of the advantages of laboratory experiments and provide some examples of the new trends in the lab experiments including the use of equipment to measure brain activity as subjects make choices.

One interesting application of this kind of experiments is the paper by Mutz and Reeves (2005). Drawing on three laboratory experiments, they examined the hypothesis that it is the manner in which disagreement is presented that discourages positive attitudes toward politics and politicians. The central manipulation in their experiments was the extent to which politicians exchanged political viewpoints in a manner that violates the typical norms governing face-to-face political conflict. The results of the experiments show that uncivil political discourse has detrimental effects on political trust. In conclusion, the format of much political television effectively promotes viewer interest, but at the expense of political trust. For Mutz and Reeves (2005, p.13)

> …extremely low levels of trust may threaten the stability of political institutions, make them to function less smoothly, and contribute to a political environment in which it is more difficult for leaders to succeed.

LIMITATIONS OF EXPERIMENTAL RESEARCH

In this last section of this review article I summarize some of the most important limitations of experiments. The specialized literature has mentioned several shortcomings of experiments. Some of these are related to the implementation of experiments and their ethical implications, others involve the population they use to generate empirical results, and finally others touch on the possibilities of experimental research replication.

First, the experimental setting is considered as artificial, limiting the external validity and generalizability of the experimental results. Based on the fact that "…politics is not a high priority for most people, trying to replicate the attention busy citizens dedicate to politics is a real concern" (Fridkin and Kenney, 2012, p.64).

In addition, experiments are often conducted with students as subjects, raising questions about whether the results can be generalized to more diverse popula-

tion. Sears (1986) has expressed this concern in great detail. He suggests that social psychology (and empirical research by default) has created biases because of its heavy dependence on a very "narrow data base": undergraduate students. Sears postulates many reasons regarding the potential hazard of the use of a narrow data base. For him, undergraduates usually come from a very narrow age range and are concentrated at the upper levels of educational background. Also, they tend to have a less than fully formulated sense of self and their social and political attitudes tend to be considerably less crystallized at this stage than later in life. They also tend to be substantially more egocentric than older adults. They differ from adults in their interpersonal relationships, as well, having a stronger need for peer approval, manifested in dependency, conformity, an overidentification with peers. These needs tend to be mixed with highly unstable peer relations and especially highly unstable peer group relationships. In sum, Sears (1986) concludes that college students in the laboratory are appropriate for some purposes and not for others.

Others have pointed to the problem of ambiguity of experimental treatments. For Kinder and Palfrey (1993, pp.24-25)

> …in an ideal experiment, we put ourselves in position to conclude that the differences we observe in behavior across the experimental treatments are caused by differences in the treatments themselves. But while experimental methods enable us to draw causal inferences of exactly this form, they do nothing to reveal to us the meaning of the experimental treatments that we create. We can say, with special authority, that the treatment did it, but what, exactly, is the treatment.

Another shortcoming is associated with the limitation in the scope of experimental research. This means basically, that, some problems that are central to political science simply cannot be investigated by experimental means. For instance, democratization and regime change are two good examples of topics that definitely are not suitable for experimental research. There is no way to simulate transitions to democracy in the lab and get persuasive results.

Experimental research is often criticized on matters of external validity. Scholars have distinguished three basic forms related to this concern (Sears, 1986). First, because experimental participants ordinarily know that they are taking part in the study of something (even if they are not sure of what), this knowledge alone may induce alterations in their behavior. Second, experiments are often conducted with samples of convenience, leading to skepticism over whether experimental results can be generalized safely to the populations of real interest. Third, experimental results are always subject to the charge that they depend on independent variables creation.

Finally, the optimistic point of view postulates

> …that although the shortcomings of experimentation are real and cannot be avoided entirely, do not constitute anything like insuperable obstacles. Indeed, the creative and energetic experimenter can even, on occasion, convert these apparent liabilities into strengths (Kinder and Palfrey, 1993, p.24).

Furthermore, the proliferation of experimental designs in the discipline increases concerns regarding ethical dilemmas. The ethical guidelines for experiments in social sciences are similar those applied in other disciplines. Three ethical principals have been identified as crucial for conducting experimental research: subjects' explicit consent, volunteerism, and anonymity. National and international agencies have implemented stricter regulations to prevent abuses and misapplications. Past and recent mishandle of experiments (like the Montana's case in late 2014) have attracted individuals' and media attention raising flags on how experiments are executed.

In conclusion, if we examine, retrospectively, the impacts of experimental political science research, the general view we get is one that reveals that the growing interest in experimentation reflects the increasing value that the discipline places on causal inferences and empirically guided theoretical refinement. In synthesis, as stated by Druckman et al. (2011, p.3)

> …experiments guide theoretical development by providing a mean for pinpointing the effects of institutional rules, preference configurations, and other contextual factors that might be difficult (or impossible) to assess using other forms of inference.

Even though, as stated by Kinder and Palfrey (1993, p.1) "…turning to experiments will hardly cure political science of all its troubles, experiments must supplement, not replace, traditional empirical methods".

REFERENCES

Campbell, D. T., Stanley, J. C., and Gage, N. L. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Converse, P. (1964). The Nature of Belief System in Mass Public. In D. E. Apter, *Ideology and Discontent.* London, England: Free Press of Glencoe.

Dickson, E. S., Hafer, C., and Landa, D. (2008). Cognition and strategy: a deliberation experiment. *The Journal of Politics,* 70(04), 974-989.

Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. (2011). *Cambridge*

handbook of experimental political science. Cambridge University Press.

Fridkin, K., and Kenney, P. (2012). Laboratory Experiments in American Political Behavior. In J. E. Leighley (Ed.), *The Oxford handbook of American elections and political behavior*. New York: Oxford University Press.

Gerber, A. S., and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 653-663.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and vote turnout: Evidence from a large-scale field experiment. *American Political Science Review,* 102(1), 33.

Green, D., Davenport, T. C., and Gerber, A. S. (2012). Field Experiments and Political Behavior. In J. E. Leighley (Ed.), *The Oxford Handbook of American Elections and Political Behavior* (pp. 69-81). New York: Oxford University Press.

Kinder, D. R., and Palfrey, T. R. (1993). On behalf of an experimental political science. In *Experimental foundations of political science* (pp. 1-39).

Lau, R. R., and Redlawsk, D. P. (2006). *How voters decide: Information processing in election campaigns*. Cambridge University Press.

Leighley, J. E. (Ed.) (2012). *The Oxford handbook of American elections and political behavior*. New York: Oxford University Press.

Lodge, M., Steenbergen, M. R., and Brau, S. (1995). The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review*, 309-326.

Morton, R. B. and K. C. Williams (2008). Experimentation in political science. In *The Oxford handbook of political methodology* (pp. 339-356).

Morton, R. B., and Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge University Press.

Mutz, D. C., and Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review,* 99(01), 1-15.

Nie, N. H. and Andersen, K. (1974). Mass Belief Systems Revisited: Political Change and Attitude Structure. *The Journal of Politics,* 36(3), 540-591.

Niemi, R. G., Weisberg, H. F., and Kimball, D. C. (2011). *Controversies in vot-*

ing behavior. Washington D.C.: CQ press.

Piazza, T., Sniderman, P. M. and Tetlcok, P. (1989). Analysis of the dynamics of political reasoning: A general-purpose computer-assisted methodology. *Political analysis***,** 1(1), 99-119.

Sears, D. 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.

Sullivan, J. L., Piereson, J. E., and Marcus, G. E. (1978). Ideological Constraint in the Mass Public: A Methodological Critique and Some New Findings. *American Journal of Political Science*, 22(2), 233-249.