

# Methodology Update: Randomised Controlled Trials, Structural Models and the Study of Politics†

Leonard Wantchekon<sup>a,\*</sup> and Jenny Guardado R.<sup>b</sup>

<sup>a</sup>Politics and Economics, New York University, New York, NY, USA

<sup>b</sup>Wilf Family Department of Politics, New York University, New York, NY, USA

\* Corresponding author: Leonard Wantchekon. E-mail: leonard.wantchekon@nyu.edu; lwantchekon@gmail.com

## Abstract

*This paper explores how the combined use of Randomised Controlled Trials (RCTs) and Structural Models can improve the study of politics. We posit that randomized controlled trials can benefit from the insights provided by structural models, particularly for the type of questions posed in Political Science. Although structural models have been utilized scarcely in politics, the close relationship between theory and empirics required by structural models would help solving many of the current pitfalls of RCTs in political science. For instance, this approach can alleviate concerns of external validity often associated with experimental evidence. We finally present a real political science example to illustrate the implementation of this approach.*

JEL classification: A1, C90, C91

## 1. Introduction

Randomised controlled trials<sup>1</sup> (RCTs) have seen an exponential increase in the field of political science. Although the first field experiment in

† This is a contribution to Centre for the Study of African Economies, Oxford University.

<sup>1</sup> In this paper we use the term of randomised controlled trials interchangeably with experimental research.

© The author 2011. Published by Oxford University Press on behalf of the Centre for the Study of African Economies. All rights reserved. For permissions, please email: journals.permissions@oup.com

political science dates back to the 1920s, with Gosnell (1927) looking at the effect of information and encouragement on voter turnout, it is not until the late 1990s when experiments became a mainstream approach in political science. Yet, the rise of experimental research in political science has not entailed simply imitating what is done in other fields. Rather, experimental research in political science has *adapted* methods from other disciplines to our own substantive concerns. Expectedly, RCTs in political science have exhibited some differences with other disciplines: First, political scientists have devoted greater attention to experiments in which the treatment resembles political processes instead of a single intervention affecting a political or economic outcome. Experiments analysing complex treatments such as community deliberation, bureaucratic processes and campaigning styles are substantively distinct from treatments consisting of a certain drug dosage and its effect on disease outcomes or the provision of anti-mosquito nets on the number of malaria incidents, or what we can denominate as the policy approach to RCTs. These substantive differences have thus motivated an innovative array of experimental research designs attempting to address some of the most pressing questions in the field. Quite expectedly, the surge of experimental research has prompted a debate on both the *feasibility* of such political science questions and their *ability* to improve extant theories in the field. Regarding the feasibility of these types of experiments, we posit that creative research designs are able to tackle these questions. For example, a few resourceful designs have managed to randomise the assignment of decision-making processes, the type of political campaigns and accountability mechanisms as we will show in Section 2. Second, experimental research in political science has also raised concerns about their overall theoretical contribution to political science. For instance, we notice that political scientists are generally more concerned with the mechanisms underpinning experimental results and not only with the overall effect of some intervention. Every so often it is not sufficient to establish a causal effect but it is becoming increasingly important to explain why and through which channel a specific outcome is observed. Cutting edge methodological research is now concerned with looking at the different experimental designs, which would facilitate mechanism testing (see Imai *et al.*, 2010) a step beyond merely assessing a causal relation. However, research looking to incorporate mediator variables and mechanism testing in experimental research might not be enough. As in other disciplines (mainly Economics), current experimental research in political science is now under pressure to become

increasingly motivated by theoretical models and less so oriented towards looking at aggregate causal effects. In this context, we motivate the use of structural models, an approach able to combine adequate theory formalisation and adequate data sources. Although seldom used in political science, structural models are able to bring together the strengths of RCTs' empirical results and a formal theory approach to the most pressing questions in the field. Yet, structural models are far from being a panacea for all the existing methodological concerns in the field and thus we shall emphasise the scope and limitations of this approach.

In sum, in this paper we aim to address the methodological aspects of current experimental research as well as the use of structural models for the study of the political process. To do so, Section 2 presents the most salient features of randomised controlled trials and structural models in political science. The section also discusses the role and potential contributions of both approaches to research. Section 3 examines in-depth an application of how both approaches can contribute to the causal inference and generalisation in a political science question. Section 4 concludes and outlines venues for future experimental research in political science.

## 2. Experimental research and structural models in political science

### 2.1 Experimental research in political science

Early randomised controlled trials in political science have primarily focused on studying the way in which various techniques of voter mobilisation (e.g. information) affected voter turnout. More recent work in political science have intended to cover a wide range of topics involving community deliberation processes, women leadership, corruption, conditional cash transfer programmes, campaign styles etc. Thus the main distinctive characteristic of recent RCTs in Political Science has been to assign 'treatments' comprising an institution or a decision-making process and assess its effect on some political outcome of interest (Atchade and Wantchekon, 2009); that is, subjects are assigned to treatments involving a decision-making process that allow them to select the treatment they will eventually receive. For instance, instead of assigning textbook treatments, flip charts or deworming treatments, political scientists randomise a 'deliberation process': students and parents are assigned to decide the amount of 'treatment' following different rules. Other examples include

the assignment of local councils to deliberate on the specific public good they want to enjoy (e.g. roads, water or sanitation projects) under some development project scheme. Similarly, villages are assigned to select and implement their own development projects in an effort to endogenously improve intra-community social cohesion levels (Fearon *et al.*, 2009). Experiments looking at municipalities that are assigned to undergo an auditing process of which the results are made public and then observing whether voters will punish electorally corruption or not is an example of this type of treatment. Local councils that are allocated a certain female quota, thus affecting the within council bargaining process over public goods, is another example. As noticed, although these RCTs share the same framework as in other disciplines, the topics are more wide-ranging and related to essentially political science questions (e.g. community participation, accountability mechanisms, women leadership). Evidently, the interest in such type of interventions arises naturally from the type of questions characterising political science: How can citizens render politicians accountable? What motivates voter's electoral behaviour? What is the effect of different bargaining rules on the political process? etc. Therefore, it is only expected that the field would approach these questions within the framework of RCTs.

Yet, the ability to include such complex treatments has also raised skepticism in the field. Even advocates of RCTs in Economics consider that the use of randomised experiments does bias researchers against large-scale policy experiments (e.g. free trade and industrial policy) (Banerjee, 2007). Thus, it is questionable whether many of the most relevant topics of interest in political science can actually be subjected to randomised evaluation (e.g. political-system reforms). The scale constraint imposes limits and narrows the range of questions sought to be answered thus narrowing the research agenda of those utilising this approach. Although economy-wide measures are indeed difficult to assess within the RCT framework, it is true that complex treatments can be implemented allowing for a wider range of political science questions to be addressed. These constraints on the scope of the questions to be answered with RCTs ought to be considered an opportunity to address these bigger questions of the field in more creative ways. For instance, landmark papers that have followed this type of treatment assignment are: Olken (2010), who conducted a field experiment in 49 Indonesian villages, each of which was to implement an infrastructure project as part of a development programme. The project was to be chosen from a list of projects set forth by the village as a whole. Each village followed

the same agenda-setting process to agree on the lists on which they would vote on infrastructure projects: a general project determined by the village as a whole, and a women's project. The experiment randomly assigned villages to make the final decision regarding the projects through one out of two possible political mechanisms: a representative-based meeting or through a plebiscite. Thus the treatment was clearly a decision-making process in which the outcome (public goods) is endogenously decided by the community.

Olken's study of these two political mechanisms—plebiscites and village meetings—in Indonesia illustrates how it is feasible to assign a 'political process' as a treatment to provide relevant answers in the field. Since the treatment assigned is an 'institution', the public good received and the degree of satisfaction and legitimation are clearly derived from the intervention. One purpose of the analysis is to see to what extent are different decision-making processes subject to elite capture. Elite capture would be an undesirable outcome if public good assignments do not reflect villagers' preferences but only those of the elite. Therefore, Olken first examines whether the *types* and *location* of the projects chosen through the plebiscite appear closer (or not) to the stated preferences of village elites. The author finds that plebiscites do not affect the *type* of project selected by the village as a whole. Rather, plebiscites appear to affect the *location* of the project chosen by women only: villages using the plebiscite as a decision-making rule chose projects that were located in poorer areas. The author suggests that the plebiscite enfranchised women who would previously not have participated in deciding the village project (e.g. poorer women). However, the type of project selected still appeared to be closer to the elite preferences, which reflects the degree of elite capture of the agenda-setting process. The strength of the result lies in the random allocation of the treatment, which reduces concerns of unobservable factors across villages that would undermine the results. As noted, the essence of the paper is to assign a 'set of rules' and observe its public good and citizen satisfaction outcome. The previous inability to randomly assign such institutions puts this paper in the forefront of the literature analysis of the effect of different institutions.

Other studies show how a bureaucratic process can also be considered the assignment of an institution. Ferraz and Finan (2008) make use of a natural experiment to study the effects of the disclosure of local government corruption practices on incumbents' electoral outcomes in Brazil's municipal elections. The research design takes advantage of the fact that Brazil had initiated an anti-corruption programme whereby the federal

government began to randomly select municipal governments to be audited for their use of federal funds. To promote transparency, the outcomes of these audits were then disseminated publicly to the municipality, federal prosecutors and the general media. Randomly assigning municipalities to an auditing treatment involve a sequence of actions in which a set of individuals (bureaucrats) 'decide' the level of corruption observed and then publicise it. This bureaucratic procedure is similar to the decision-making process described above once we realise that the effect of the treatment (auditing and publicising) is contingent on the identity of the auditing agency and the mechanism to publicise the results: were the result to be publicised differently, or were the audit to be conducted by another entity (e.g. journalists), the observed outcome could be very different. In addition, the series of decisions from the 'bureaucracy' indicates that at least some part of the treatment could be determined endogenously: types of reported issues, the specific content of the report, disclosure methods, even the bureaucrat himself, imply an action that could have some impact on the treatment (auditing) and subsequently the outcome. For instance, variation in the style in which the corruption report is written (e.g. in more technical terms rather than in journalistic lingo) or variation in the government agency evaluating local governments or in the quality of local media would expectedly have an effect on citizen's reaction to corruption reports. This of course does not undermine the results obtained by the authors; it just highlights the 'process-like' nature of the treatment assigned to these municipalities. Finally, Ferraz and Finan compare the electoral outcomes of mayors eligible for re-election between municipalities audited before and after the 2004 municipal elections. The authors do find that incumbents audited before the election did worse than those audited after the election, conditional on the level of corruption exposed.

Lastly, [Chattopadhyay and Duflo \(2004\)](#) study the effect of the quota system for women's political participation in local councils on the provision of public goods in India. The natural experiment occurred by the implementation of the 73rd Amendment, which required that one-third of local Council head positions be randomly reserved for women in the selected villages. To check the accuracy of the randomisation procedure, the authors verify that in the two districts studied (West Bengal and Rajasthan), all positions of chief in local village councils (Gram Panchayats, GPs) reserved for women were, in fact, occupied by females. According to the authors, the presence of female heads at the local council is expected to change the policy outcomes of the GP in these

districts: since the public goods preferences of female representatives tended to differ from those of males in these villages, the policy outcome under the intervention is expected to be different. We can consider this treatment as one involving a decision-making process to the extent that this mandated representation of women implies an endogenous selection of the treatment received: female *pradhans* (heads) choose to promote a specific type of public goods (e.g. those most favoured for women in their district or themselves) over other types. So, in this experiment instead of assigning directly those public goods preferred for women, these were rather endogenously selected, which might have an effect on the outcome of interest such as future satisfaction of the council's workings, increased female welfare, among others. Although the latter is not among the purposes of the authors, one can readily hypothesise that such effects are plausible. Effectively, Chattopadhyay and Duflo's evidence confirms that this random assignment of women to local council's leads to an increase in those public goods reflecting local women's interests: water and roads in West Bengal and water in Rajasthan. Moreover, having a woman in the local council also increased the involvement of women in GPs' affairs in West Bengal, although not in Rajasthan. As said, the experiment exemplifies a treatment involving an endogenous selection of a treatment at the level of GPs in India.

As stated above, the growth of RCTs implementing more institution-like 'treatments' in political science has been driven by the substantive concerns of the discipline and of course a better knowledge and familiarity with the implementation of RCTs. However, it should be noticed that the assignment of the latter process type of treatments has also raised issues about the actual possibility of isolating the effect of institutions from the treatment selected endogenously (Atchade and Wantchekon, 2009): if the outcome is determined by the decision-making procedure, how to distinguish the effect of the rules (institution) from the actual decision made on subsequent outcomes obtained? This is most clearly highlighted in Olken (2010). One of Olken's (2010) main conclusions is that citizen's satisfaction is higher under plebiscites than under representative meetings for both general projects and women's projects. Yet, since the location of women's projects differed under plebiscites and village meetings, it is unclear whether the satisfaction with democracy is driven by the institution or the policy outcomes: What is the increase in satisfaction derived from the exercise of democracy (plebiscites) and that derived from the good provided? For Ferraz and Finan (2008) the effect of the rules (e.g. revising mayor's finances) cannot be disentangled from the

fact that the audits were also publicised. Although not relevant for the purpose of the paper, subsequent research might be interested in the separate effect. Similarly, in [Chattopadhyay and Duflo \(2004\)](#), it is possible to look at the effect of the reservation policy on other social outcomes (e.g. larger social acceptance of women, among others) which would then need to distinguish the effect of the rule from that of the public good obtained.

In addition, one of the arguments we emphasise is that the ability to distinguish the effect of institutions from that of policy outcomes on subsequent indicators can be improved by adopting a more theoretical approach while designing experimental research in general; that is, once we theoretically characterise the treatment as a political process with a number of different intervening variables, we can best distinguish how these jointly affect the observed outcome. This is of course a step further than simply obtaining a causal effect between treatment X and some outcome Y. Since we know that the treatment X is composed of different elements, which can themselves have a separate effect on Y, we need to seriously model these intervening variables accompanied by the appropriate empirical tests accounting for these intermediate effects. This is true for those experiments in which we suspect there is a presence of mediator variables and those cases in which we are explicitly including intermediate outcomes as is the case of 'institutional' treatments as the previous examples illustrate.

Despite the criticisms of the narrowness of the approach, many political science experiments do follow the 'one-intervention' framework. For instance, [Wantchekon \(2003\)](#) implements a 'clientelist' treatment to a sample of randomly selected villages during Benin's presidential campaign of 2001 and then looks at the political participation outcome. Wantchekon worked directly with presidential candidates to randomly select villages to be exposed to purely clientelist or purely public policy platforms. Unlike the articles mentioned above, the treatment in [Wantchekon \(2003\)](#) did not involve a decision-making process. Instead, villages assigned to different messages were expected to exhibit different political participation effects. While the average effect of the non-clientelist campaign messages was to reduce support, those villages with a clientelist appeal exhibited a positive effect on the support for the party endorsing it. These effects were stronger among a subset of characteristics: incumbent candidates and female voters. If gender and incumbency are relevant variables to be considered when analysing the appeal of campaign messages, an appropriate theoretical modelling of the mechanisms through which voters are



appealed to by different campaigns would need to account for these observed effects.

Another example of a single intervention experiment is that of Gerber *et al.* (2007) looking at the effect of information on voter turnout. In 2005 the authors conducted an experiment during the Virginia gubernatorial election. The purpose of the experiment was to analyse how *partisan* information affected voter behaviour. The subjects were selected before the election in two waves from lists of registered voters and a consumer database list. In the first wave, individuals were surveyed and asked about their newspaper habits and sources as well as other control covariates. Gerber, Kaplan and Bergan acknowledge the fact that Washington, D.C. has two main newspapers, one of which is perceived as a liberal newspaper and the other as a conservative one. Thus the treatment consisted in randomly assigning subjects to a free 1-month subscription to the liberal newspaper, or a free 1-month subscription to the conservative newspaper, and a group that received neither offer. The main result is that those individuals assigned to 'liberal' newspapers were more likely to vote Democrat than those not assigned to any newspaper. However, if we observe that this effect is stronger among certain types of 'liberals' or even if we suspect other intermediate variables, an appropriate theoretical approach should account for it as stated above.

This leads us to the *second most salient feature* of RCTs in political science, which is the growing attention to the role of causal mechanisms on the outcome studied. Very often the theory motivating RCTs only refer to intuition and not to formal models specifying each intervening variable. Yet, due to the substantive concerns of the field, it often does not suffice to assess the impact of a policy intervention. Rather, political scientists mostly care about *how intervening variables produce the observed outcome*; that is, political scientists are interested in tracing the specific mechanisms through which the effect is achieved. For instance, stating that democracy itself leads to higher citizen satisfaction requires not only isolating the increase in satisfaction due to the public goods democratically provided, but also a theory of why satisfaction is greater when citizens cast a vote in a general election. Is satisfaction greater due to the expressive value of the vote? Would citizen's satisfaction be the same with other mechanisms of opinion consulting (e.g. survey polls) or under different voting rules? The answer to these questions has implications for policy making and the theory of democracy in general. Results derived from RCTs that are loosely guided by theory are often deemed inconclusive in terms of their theoretical contribution: Stating that X has an effect on Y but not

*why* we observe such an effect is increasingly insufficient to have an impact on theory building. Unless the experiment is explicitly derived from a theory, the mere unravelling of causal relationships is expected to have little impact on theory. Of course, these experiments may be useful for fact discovery and hypothesis building. Yet, theorising about the mechanisms leading to an observed effect can significantly improve the ability to apply these estimated ‘effects’ in settings never previously implemented (external validity), which is not always plausible when merely engaging in fact discovery. Moreover, it is not until recently when greater attention was paid to causal mechanism and research designs able to account for it. For instance, Imai *et al* (2010) looks at experimental designs in which the causal effect of the treatment can be estimated in the presence of mediator variables. Thus, different experimental designs are able to test complex statements once these are adequately accounted for.

An example of the care given to the role of mediator variables is shown in Bertrand and Mullainathan (2004) (in Imai *et al.*, 2010) who conducted a randomised field experiment sending fictitious resumes manipulating the perceived race and estimating the call-back rates. In this experiment, the fictitious ‘qualifications’ of the applications would clearly work as a mediator variable: different qualifications might modify the effect of the perceived race. For example, discriminated ethnic groups with high levels of qualifications might have higher call-back rates than non-discriminated groups with the same level of qualifications under the premise that higher qualifications in an adverse environment is a strong signal of quality. Therefore, such qualifications were correctly held constant throughout the experiment so as to not introduce such heterogeneity and clearly account for the perceived race effect. However, it is not always possible to manipulate to such a degree the intervening mediator variable. For instance, in Olken (2010), a mediator variable is the public good decided via plebiscite or the representative-based meeting, which is not manipulated by the author. As said above, if we think that the type and amount of public goods observed matters, then it should account for as an intervening variable as well. Similarly, in Ferraz and Finnan (2008), one of the intervening variables is the audit report. The report is a product of the auditing process which may have an effect on the electoral support of voters. As noted, the report is process specific; it could have had a different effect if the report were generated through an investigative journalism process, for example. Thus, addressing the so-called black-box view of causality is one of the priorities of the field, thus prompting an increase

in structural modelling in political science as we shall discuss in the next section.

## 2.2 Structural models in political science

As in the case of RCTs, the distinctive feature of structural models in political science lies in the substantive topics they address as well as in the attention paid to the theoretical assumptions between the causes and effects analysed. Although scarce, structural models in political science have been present at least since the 1970s in the sub field of international relations (Morton and Williams, 2010). More recently, structural models in political science have been mostly applied in voting and turnout topics, institutional design and deliberation models. Papers on the latter topics exhibit many parallels in the way economists approach certain topics (e.g. in industrial organisation) and thus are published in Economics journals even if addressing a substantive political science topic.

An example is Diermeier *et al.* (2003), paper that develops a bargaining model of government formation in parliamentary democracies. Specifically, the authors address the effect of different constitutional rules of government formation and termination on the duration and size of the government coalitions in parliamentary democracies. Unlike other empirical studies in the same topic, Diermeier *et al.* (2003) provide a model of government formation able to account for the empirical regularities in the data as well as the hypothetical effect a change in rules would have on the government formation process. Such ability to generalise derives both from the formal theoretical approach taken to derive the testable propositions and subsequently estimate the empirical models closely following the latter. This paper is one of the few in political science using a structural modelling approach. Another example of the structural approach in political science is that of Coate and Conlin (2004) who develop a model of voter turnout based on the costs of voting against the benefits of a change in current policy. Using the theoretical predictions, the authors assess the models' performance by using real turnout data for policy referenda in the USA. The authors then structurally estimate the models to fit the empirical data available.

Both papers highlight the elements as comprising structural models, and in the case of Diermeier *et al.* (2003) we can see the following: First, prior theory shows the cause–effect relationship between specific institutional features of parliamentary democracies (exogenous variables) on the formation and dissolution of coalition governments (endogenous variables).

Second, the testable propositions derived are also specific about the restrictions on the statistical model to be estimated: which variables are time invariant, which are considered random variables, which is the functional form of the joint distribution of the latter variables and so on. Yet, some issues arise from the latter: first, although that these restrictions on the statistical model parameters are derived directly from the theory, we often need to assume that these conditions are fulfilled in the data which may not be self-evident. As noticed, the latter concerns are unavoidable; these can be alleviated with transparent and well-substantiated justifications of the functional forms assumed.

Third, Diermeier *et al.* (2003) use observational data on government size, duration and constitutional rules from observational data. Yet, experimental data can be used to alleviate current problems of identification associated with the estimation of structural models (Imbens, 2009). For instance, it is often 'difficult to identify and estimate the full array of structural parameters' (Heckman and Urzua, 2009, p. 2). Finally, once a model is estimated, it is hard to assess the validity of a certain structural model. Yet, these concerns can be reduced when comparing the performance of the model with actual policy changes, or in different environments. The ability to evaluate the goodness of fit of different structural models improves their prospects of external validity. For instance, once we reach the right specification, it can be readily tested with additional data.

The use of structural modelling in political science is still not so widespread in the field of political science, yet we contend that its use can increase due to: (i) the similarity between many economic situations and political science phenomena of interest. For instance, there is a resemblance between the study of auctions and that of elections. Also, structural models on firm competition can resemble models of political competition in the field (Morton and Williams, 2010), (ii) the *positive feature* of structural models that there is a larger role for theory to accurately characterise (including assumptions) the studied relationship places an advantage over studies looking at 'atheoretical' RCTs, (iii) although problems of identification, specification and estimation can be very complicated when using structural models (Cameron and Trivedi, 2005); the use of experimental data can alleviate concerns of the identification of structural model parameters as shown in landmark papers in economics (Imbens, 2009). For instance, Duflo *et al.* (2010) looks at the monitoring and financial incentives that would reduce teachers' absenteeism in Indian schools. As a treatment they assign every village school to a camera assignment (monitoring) accompanied with a payment (incentive) if complied with

or exceeded the allocated work hours. First, they look at the obtained absence rates from random checks as well as from what is revealed by the pictures. The authors show an immediate and long-lasting increase in teacher attendance. Moreover, the authors then propose a dynamic model of labour supply to incorporate the teacher's incentives varying over the month. Using this model, they estimate it structurally to predict the amount of expected days of work. Finally, they compared the predicted result with the actual experimental data. This paper is an example of the possible collaboration between experimental data generated via RCTs and a structural modelling approach. In the next section, we will discuss how these research approaches have been utilised in the field of political science and discuss their influence.

### 2.3 RCTs or structural models in political science?

So far, we have described and provided some examples of political science research conducting experimental research or using structural models. Yet, which approach should prevail in the field of political science? The answer is that both types of research should be conducted in political science considering each approach can learn from one another's weaknesses and strengths. For instance, *RCTs* have motivated the field to think about causality and parameter identification in more rigorous terms. In turn, *Structural Models* push towards making explicit the theoretical mechanism underpinning current experiments used in political science. In general, a combined use of RCTs and structural models would improve the political science literature in three aspects.

First, a combined use will help to produce a more theory-oriented research directed towards unveiling causal mechanisms behind political phenomena. RCTs have to be clearly founded in theory: following a specific model, researchers will make a better use of evidence across apparently disparate studies. For instance, *Imai et al. (2010)* constitute a clear attempt to improve the contribution of RCTs to try and directly test a theory and not only an expression of it. Therefore, RCTs would need to improve their design and theory orientation to have a larger impact on theory construction. In this sense, structural models have the advantage of seriously considering theory in their estimation strategy. In fact, the most important part in the specification of the structural model is the prior (theory) ordering of variables into causes and effects (*Cameron and Trivedi, 2005*).

Second, a more theory-oriented research would also alleviate concerns of external validity, which is an on-going debate in political science and in

other fields. In this sense, an advantage of structural modelling is that by making explicit the mechanisms behind a presumed effect, it facilitates the extrapolation of results to new data. This is consistent with what Martel Garcia and Wantchekon (2010) call the *analytical approach* of external validity. The authors see the problem of generalisability as ‘...intrinsically theoretical, in that theories about causal mechanisms, constructs, and selection are what allow us to generalize beyond sampling particulars in individual cases’. The analytical approach requires researchers to put at the centre of the experiment the mechanisms that link a causal variable to an outcome. By being explicit about mechanisms, researchers can design RCTs’ testing theoretically informed hypotheses. Specifically, researchers would have to: (i) theoretically assess to what extent the assumptions of the experimental units are similar to the relevant aspects of the new subjects, (ii) look at the out of sample predicted cause and effect relationship, which is greatly improved once all relevant moderators or additional covariates are included in the model (well-specified model). Broadly speaking, these two issues are also applicable to the estimation of structural models, and thus the estimation procedure can be considered fairly.

A second approach to alleviate concerns of external validity has been the accumulation of experimental evidence results across different settings without varying the treatment assignment. The so-called robustness approach underpins much of the motivation behind replicating RCTs in different contexts, which are seen useful to uncover empirical regularities and developing theory (Gerber and Green, 2002). Yet, when the components of treatments vary (even slightly) across studies, *knowledge does not accumulate across studies* (Heckman and Vytlačil, 2007). Moreover, it is not clear how many replications would be necessary to establish a general result and which part of the treatment would need to vary between these studies. Thus, we advocate a more analytical approach to establish the external validity of certain experimental results, which necessarily entails closer ties to theory and pushes the field towards adopting a view closer to what structural models actually advocate.

Finally, the incorporation of structural models would allow for more complex treatments which comprise a great part of most interesting questions in political science; that is, the most relevant ‘treatments’ in political science are often complex packages of interventions. In this sense, the field of political science can benefit from ‘unbundling the treatment’: tracing the specific mechanisms through which the effect is achieved. At the moment, RCTs in political science are limited in their ability to ‘unbundle’ these

treatments, mostly because we are unable to trace how the same mechanisms can affect more than one outcome at the same time even if we presume such effects exist.

That said, RCTs should exhibit a *leading role* among the methodological choices of political scientists for the following reasons. First, as mentioned before, RCTs are increasingly becoming more relevant to address substantive questions in political science. Even with the limitations described above, researchers have been increasingly able to find creative ways to design treatments that resemble the variables of interest in political science, even if highly complex. Second, theory formalisation is a precondition for structural modelling that is not yet fully met in many areas of political science. Even in areas with highly formalised theory such as bargaining models of government formation, the ‘...theoretical contributions typically aim at providing tractable models that explain some of these facts [of government formation], but are in general not suitable for empirical analysis’ (Diermeier *et al.*, 2003, p. 29). Finally, event uniqueness and complexity of the phenomena studied in political science make difficult the implementation of structural models due to data availability on the foundations of these events. For instance, structural models are well suited to analyse very detailed descriptions of political (or economic) behaviour; to the extent that the existing accounts depart from these stringent conditions, structural modelling cannot be applied.

In sum, structural models can contribute to the study of politics, yet the RCT’s have been and will continue to have this leading role in the discipline.

### **3. RCTs and structural models: a political science application**

To illustrate the points discussed above, we present a hypothetical political science problem: how to assess the effect of campaign platforms on political participation and the demand for different public goods? This question was actually addressed in an experiment conducted by L. Wantchekon in 2006. Specifically, the experiment studies the effect of different types of campaigning strategies on turnout, voting and voter’s knowledge in a certain constituency. The treatment consisted in assigning candidates to use village town hall meetings to promote public debates around specific policy platforms to spread their campaign message. The control group comprises those villages in which pre-designed campaign messages were delivered through public rallies and no interaction between candidates

and voters occurs. The relevant comparison is then between those villages in which campaigns allowed for endogenous selection of policies through public deliberation and those in which policies were pre-fixed or ‘exogenously’ assigned. Since both the treatment (town hall meetings) and the outcome (turnout) are clearly defined, the most straightforward RCT design would suggest to randomly assign half of the villages to conduct a town hall meeting while the other half should be subjected to normal campaigning strategies (e.g. rallies). The estimate of interest would be the Average Treatment Effect (ATE), which is the difference in the average turnout, voting preferences and political knowledge among both groups. The actual experiment took place in Benin involving five candidates running in the first round of the 2006 presidential elections. This experiment is a follow-up from a previous experiment, which finds that promises of particularistic goods are a more effective electoral campaign strategy than platforms focused on general-use public goods (Wantchekon, 2003). If the lack of support for non-clientelist platforms is due to the vagueness used in the campaign messages, is it possible that controlling for this factor, programmatic politics can become more attractive to voters? The author contends that effectively different types of campaigning can affect measures of political participation such as voting, turnout, and the voter’s knowledge of the campaign and the issues at stake. Experimental results suggest that broad-based platforms can be effective in generating electoral support when they are specific and communicated to voters through town hall meetings (Wantchekon, 2009). The findings suggest that: (i) the treatment has a positive and significant effect on policy information, (ii) turnout was significantly higher in treatment villages than in control villages and (iii) the use of information on public policy and dissemination is at least as powerful as vote buying in terms of garnering votes for the candidate using town hall meetings.

Now, suppose that we formally derived theory showing that town hall meetings lead to ‘specific’ campaign messages. If we suspect that ‘message’ is a mediator variable intervening in the observed causal effect, this concern can be addressed with a proper RCT design. Following Imai *et al.* (2010), a second randomised experiment can be conducted to assess the causal mechanism in cases where the messages cannot be directly manipulated. For instance, we can first randomly split the sampled villages into two groups and assign one group to the town hall meeting treatment as before. To the control group, we can randomise both the town hall meeting treatment as well as the *encouragement given* to certain candidates to campaign on certain messages. The effect of the campaign type will then



be the comparison between the first group, and those who were encouraged to convey a certain message type. This procedure will allow us to account for the mediating role of messages, which will be different from that of the campaign style. Therefore, it is possible to only use RCTs to address this question. Yet, criticisms of external validity are inevitable: how applicable is this mechanism to other settings? Even if the mechanisms are suspected to go through the hypothesised variable, what is the functional form of this relationship? How can baseline covariates affect the observed outcome? These and other questions can be better answered with the incorporation of structural models. Therefore, we return to our formally derived model, in which we show that some types of messages ( $m$ ) cause larger turnout and support ( $y$ ) than others. Once this relationship is established, suppose that certain covariates (e.g. size of the town hall) are likely to affect the message issued from which we have collected data on, say, vector  $\mathbf{z}$ . From the cause-and-effect relationships given by the theoretical model, we *assume* a functional form of the relationship<sup>2</sup> derived from the theory as well as the restrictions on the parameters of the model. Further, we assume that there is a random shock term ( $\varepsilon$ ) reflecting unobserved village characteristics that may affect the type of message ( $m$ ). The task is to explain the levels of turnout as a function of the vector of variables  $\mathbf{z}$  and the random disturbance  $\varepsilon$ . Thus, for the observed  $\mathbf{y}$ , variables  $\mathbf{z}$  and disturbance ( $\varepsilon$ ), it must be that

$$g^{(\text{turnout } y_i, z_i, \varepsilon_i | \mu) = 0},$$

where  $g$  is a known function (assumed) and  $\mu$  are the structural parameters. The purpose of the estimation of  $g$  is to infer the elements of  $\mu$ , for instance, to generalise in distinct settings. However, note that the set of variables  $\mathbf{z}$  are treated as exogenous, which is too strong of an assumption. For instance, maybe only certain sizes of town hall meetings affect the message, and conversely, more interesting ‘messages’ may attract a larger crowd. Thus, the RCT can help justify the exogeneity of the intervening variables; that is, if the RCTs fixed the size of the meeting, the effect of this covariate on the outcome can be either maintained constant or assessed through the structural estimation of the model and would no longer cast doubt on the identification of the ‘message’ parameter. In this sense, RCTs and structural models can jointly benefit from each other’s strengths and alleviate their possible weaknesses.

<sup>2</sup> Could be a parametric, semi-parametric or non-parametric specification.

#### 4. Conclusions

The rise of experiments as one of the most prominent empirical strategies in political science has brought new insights into the study of politics, yet this rise has also been accompanied with increased criticisms. Recent experimental research has covered a wide range of topics such as women leadership, corruption, conditional cash transfer programmes, clientelist and programmatic politics. Yet, in contrast to other disciplines, it is often not enough among political scientists to establish a theoretically relevant finding. Rather, due to the substantive topics of interest, the objective is to uncover the causal mechanisms underpinning political behaviour.

Although this interest for causal mechanisms would naturally suggest the use of structural models, the fact is that the use of structural models in political science is limited. For instance, the levels of theoretical formalisation are still incipient in many sub fields of political science. Second, the nature of many political science phenomena makes difficult data collection, assessment and evaluation. Yet, it is possible to adopt some of the structural model features to improve current RCT designs. For instance, certain experimental designs alleviate concerns regarding theory testing, which is a strong point of structural models. Moreover, the stress put on theory can help improve the external validity of current RCTs.

Finally, a way to use both approaches is shown in a political science application. In this example, it is possible to identify the effect of the type of campaign on political participation using experimental data. Yet, a structural model correctly specified would more straightforwardly allow for the parameters estimated in these villages to naturally extrapolate to other settings. A way to improve the estimation of structural models would be to incorporate the data collected through the experiment. In the presence of both types of data, identification and external validity concerns are greatly alleviated. Thus, despite the challenges, experiments are a leading research tool with the potential to make more substantial contributions to the study of political science.

#### References

- Atchade, Y. and L. Wantchekon (2009) 'Randomized Evaluation of Institutions: Theory with Applications to Voting and Deliberation Experiments', Working Paper, New York University.
- Banerjee, A.A. (2007) *Making Aid Work*. Cambridge, Massachusetts: MIT Press.

- Bertrand, M. and S. Mullainathan (2004) 'Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination', Mimeo.
- Cameron, C.A. and P.K. Trivedi (2005) *Microeconometrics*. New York: Cambridge University Press.
- Chattopadhyay, R. and E. Duflo (2004) 'Women as Policymakers: Evidence from a Randomized Policy Experiment in India', *Econometrica*, 72: 1409–43.
- Coate, S. and M. Conlin (2004) 'A Group Rule: Utilitarian Approach to Voter Turnout: Theory and Evidence', *The American Economic Review*, 94 (5): 1476–504.
- Diermeier, D., H. Eraslan and A. Merlo (2003) 'A Structural Model of Government Formation', *Econometrica*, 71 (1): 27–70.
- Duflo, E., R. Hanna and S. Ryan (2010) 'Incentives Work: Getting Teachers to Come to School', Mimeo, MIT.
- Fearon, J., M. Humphreys and J. Weinstein (2009) 'Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia', *American Economic Review: Papers & Proceedings*, 99 (2): 287–91.
- Ferraz, C. and F. Finan (2008) 'Exposing Corrupt Politicians: The Effect of Brazil's Publicly Released Audits on Electoral Outcomes', *Quarterly Journal of Economics*, 123: 703–45.
- Gerber, A. and D. Green (2002) 'Reclaiming the experimental tradition in political science', in I. Katznelson and H.V. Milner (eds), *Political Science: State of the Discipline*. New York: W.W. Norton, pp. 805–32.
- Gerber, A., D. Karlan and D. Bergan (2007) 'Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions', Mimeo, Yale University.
- Gosnell, H. (1927) *Getting Out the Vote: An Experiment in the Stimulation of Voting*. Chicago: University of Chicago Press.
- Heckman, J.J. and E.J. Vytlacil (2007) 'Econometric evaluation of social programs, part I: causal models, structural models, and econometric policy evaluation', in J.J. Heckman and E.E. Leamer (eds), *Handbook of Econometrics*. New York: Elsevier, pp. 4779–874.
- Heckman, J. and S. Uruzua (2009) 'Comparing IV with Structural Models: What Simple IV can and Cannot Identify', Working Paper 14706, National Bureau of Economic Research.
- Imbens, G.W. (2009) 'Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)', Mimeo, Harvard University.
- Imai, K., D. Tingley and T. Yamamoto (2010) 'Experimental Designs for Identifying Causal Mechanisms', Mimeo, Princeton University.
- Martel Garcia, F. and L. Wantchekon (2010) 'Theory, External Validity, and Experimental Inference: Some Conjectures', *The ANNALS of the American Academy of Political and Social Science*, 628: 132–47.

- Morton, R. and K.C. Williams (2010) *Experimental Political Science and the Study of Causality*. New York: Cambridge University Press.
- Olken, B. (2010) 'Direct Democracy and Local Public Goods, Evidence from a Field Experiment in Indonesia', *American Political Science Review*, 104 (2): 243–67.
- Wantchekon, L (2003) 'Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin', *World Politics*, 55: 399–422.
- Wantchekon, L. (2009) 'Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin', Working Paper, New York University.