

Review Article

*Damien Bol** **Putting Politics in the Lab: A Review of Lab Experiments in Political Science**

Experiments are now common in political science. They are an excellent methodological tool to estimate the causal effect of a treatment on an outcome. In this article, I review the use of lab experiments in political science. After a brief report on their popularity and advantages, I distinguish two ideal-types (economics-based and psychology-based) and outline the main lines of division between them. In the final section, I discuss the main challenges that lab experimentalists are facing today.

Keywords: experiment; lab; method; political science

The founding fathers of political science were convinced that they could not use experiments in their research because it was impractical and unethical (Lowell 1910). Whereas natural scientists could easily manipulate non-living elements without fearing the negative consequences the research would have on their objects of investigation, political scientists (or more generally social scientists) considered that they could not and should not conduct experiments with human subjects. Consequently, for decades, the most influential political scientists recommended the use of the comparative method as a substitute for experiments (Lijphart 1971; Przeworski and Teune 1970; Ragin 1989). The comparison of different-yet-similar people, countries, regions, time periods and so on was presented as the best tool to unravel the causes of an outcome. This idea is at the root of many political science methods such as qualitative case studies or quantitative regression analysis. I call them ‘observational methods’ hereafter.

Yet, since 1990, a substantial number of papers using experiments have been published in the most prominent political science journals (Morton and Williams 2010). In this article, I offer a review of the use of lab experiments in political science. In the following section,

* Damien Bol is Assistant Professor in the Department of Political Economy at King’s College London. Contact email: damien.bol@kcl.ac.uk.

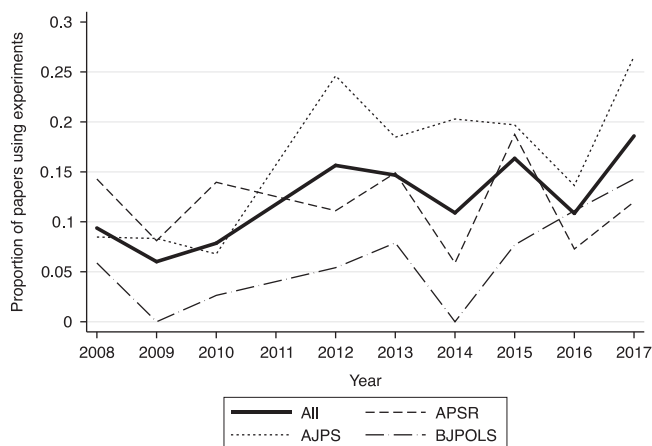
I describe the advantages and limits of experiments compared to observational methods. Further, I offer a typology of lab experiments and argue that they can be classified into two groups: economics and psychology experiments. I outline the lines of division between the two groups: (1) the outcome of interest (behaviour vs attitude); (2) whether the responses of subjects to the experiment are incentivized with money or not; (3) the degree to which the experimental design is an abstraction of the situation studied (or a realistic reproduction of this situation); and (4) whether the research focuses on subjects individually or on their interaction with others. In the final section, I discuss some of the main challenges that lab experimentalists are facing: the use of convenient samples, ethics (including deception) and reproducibility.

THE POPULARITY OF LAB EXPERIMENTS

In this section, I show that experiments are increasingly popular in political science and that they have been used to study a wide range of political topics. I define ‘experiments’ as studies for which the researcher seeks to estimate the causal effect of a treatment on an outcome (also known as the ‘treatment effect’) after randomly assigning the treatment to some of the cases, but not all of them, and by comparing the value of the outcome in the cases that received the treatment to those that did not. As I explain below, this method allows the researcher to estimate the average treatment effect for the outcome. It is important to note that I adopt a conservative definition of the method, as I exclude natural and quasi-experiments for which the researcher does not assign the treatment herself, but relies on the random (or ‘as good as random’) variations of treatments as they naturally occur in reality. My definition of experiments includes lab, survey and field experiments.

I reviewed all the papers that contained the term ‘experiment’ in their title or abstract that were published in the last 10 years in three of the flagship journals of the discipline. To avoid subfield biases, I selected generalist journals published in the US and Europe: the *American Political Science Review* (APSR), the *American Journal of Political Science* (AJPS) and the *British Journal of Political Science* (BJPOLS). In total, I reviewed 1,473 papers, of which 176 are experimental (sometimes experiments are used in combination with other

Figure 1
Experiments in Political Science Journals



methods), i.e. 11 per cent. Figure 1 reports the evolution of this proportion from 2008 to 2017. Unsurprisingly, we observe a positive trend: combining all three journals (the line ‘all’ in the figure), the proportion of published experiments goes from 10 per cent in 2008 to almost 20 per cent in 2017. Also, Figure 1 reveals that the AJPS is the journal that has published the most experiments, followed by the APSR. However, BJPOLS is catching up rapidly, as almost 15 per cent of the papers published in 2017 use experiments.

As mentioned above, this article focuses on a specific type of experiments: lab experiments. Experiments are usually classified depending on their location (Morton and Williams 2010). Following this criterion, they can be in the lab or in the field (Bol 2019). By ‘in the lab’ I mean experiments for which the subjects come to a place that is maintained by the researcher (or a research assistant), typically in the campus of a university, to participate in the experiment.¹ In the lab, the researcher usually seeks to recreate a situation that resembles a real-life one, and then randomly assigns a treatment to some subjects in order to observe their reaction. Consequently, the researcher has maximal control over the data collection process, from the treatment assignment to the measurement of the outcome. She can be sure that the treatment is correctly assigned and that the outcome is accurately measured. This differs from experiments in the field (including both field and survey experiments), which are

conducted outside the lab. With experiments in the field, the researcher has less control over the treatment assignment because she is not there to make sure that the subjects noticed the treatment.

A substantial number of the papers that I reviewed used lab experiments (around 40, or nearly a quarter of all papers using experiments). They cover a broad range of topics in all subfields of political science. For example, they are frequently used in comparative politics: John McCauley (2014) used the method to study identity politics in ethnically divided African countries; Michael Gilligan, Benjamin Pasquale and Cyrus Samii (2014) studied prosocial behaviours in communities affected by violent insurgencies in Nepal; Claire Adida, David Laitin and Marie-Ann Valfort (2016) studied discrimination against Muslims in France. Lab experiments have also been used in international relations: Scott Gartner (2008) used them to study the effect of casualties on public support for war; Lesley Terris and Orit Tykocinsky (2016) studied the process of international negotiations between government leaders. Other studies have used the method to study big questions that transcend subfields such as political legitimacy in a Weberian framework (Dickson et al. 2015), or whether people are self-interested and/or altruistic in collective decision-making processes (Sauermann and Kaiser 2010). In this review, I mostly draw from studies about elections in established democracies. However, lab experiments can be used for all sorts of political science topics. As with qualitative and quantitative observational methods, the imagination of the researcher is the only limit.

THE ADVANTAGES AND LIMITS OF EXPERIMENTS

There are two main advantages to experimental methods. The first is that they have the potential to generate more valid answers to causal questions than observational methods. The second is that they can do so on the basis of some very simple statistical tests.

Most political science studies aim to identify the causal relationship between a variable X and an outcome Y (King et al. 1994). Arguably, experimental methods are better equipped to arrive at a valid answer to these studies than observational methods. This is because the researcher manipulates the variable X herself, and proceeds to random assignment.²

With observational methods, there is no manipulation or random assignment. The researcher observes what happens in real life, and the variable X is naturally present for some of the subjects, and absent for others. The problem is that the variable X is itself the result of multiple causes. A subject has more chances to have X depending on other variables. Therefore, the group of subjects for which X is present is not perfectly comparable to the group of subjects for which X is absent.

What an experimentalist does to address this issue is randomly divide the subjects into two groups, and assign the variable X to only one of them. The group where the variable X is present is called the 'treatment group', and the group where the variable X is absent is called the 'control group'. The advantage of this procedure is that the two groups are perfectly comparable. There is no other variable that can affect the probability of a subject having X because it is the researcher herself that divides the subjects into two groups and assigns the variable X to only one of them.

Random assignment solves two specific problems. First, it solves the problem of reverse causality. Imagine that one of the other variables that cause X is Y itself. With observational methods, the researcher observes an association between X and Y. For example, she observes that the value of Y is larger when X is present than when X is absent. It is not sufficient to conclude that X causes Y, because it is equally likely that it is Y that causes X. The only thing that she knows is that the two are associated, but the causality can go both ways. With experiments, the researcher can identify what is the cause and what the effect because she manipulates X herself. It is therefore impossible that Y causes X in her design.

Secondly, random assignment solves the problem of the omitted variable bias. Imagine that the other variables that cause X are also associated with Y. With observational methods, the researcher would need to control for all these extra variables in her analysis. If she fails to do so, her estimates of the causal effect of X on Y would be incorrect. The other variables that exist in the reality but are not included in the analysis would confound the estimates. This is virtually impossible in a discipline such as political science because the topics are usually the result of complex social interactions. There is a countless number of variables that the researcher would need to include in her analysis. With experiments, this problem disappears because it is the researcher herself who decides which subjects

go into the treatment group, and which subjects go into the control group. The other variables cannot influence the random assignment.

To illustrate the advantage of experimental over observational methods for causality, imagine a study for which the researcher wants to estimate the effect of a proportional electoral system (variable X) on electoral turnout (outcome Y). Imagine the researcher observes that in real life turnout is high in countries in which there is a proportional electoral system and low in countries with other types of election systems. She might be tempted to conclude that the presence of proportional representation causes a high turnout. However, she cannot be certain that this the right direction of causality. Evidence shows that political elites design the electoral system to fit the social structure of a society (Boix 1999). There is a possibility that turnout was already high when they adopted a proportional system.

Also, there are other variables that affect turnout, and some of these variables also probably affect the probability that a country has a proportional system in the first place. For example, the number of social cleavages is likely to affect both: the population of multi-cleavage societies could be more politically engaged due to the presence of several political divides in the population (Amorim Neto and Cox 1997). The researcher would need to include all the variables that could potentially affect the electoral system and turnout in her analysis. However, she could never be sure that she had included all of them. For these two reasons, observational methods cannot identify the causal effect of a proportional system on turnout. The only solution would be to run an experiment which randomly divides the countries into a treatment group and a control group, and apply a proportional system in only one of them.

There is also a second non-negligible advantage of experiments: they do not necessarily require the mobilization of complex statistical techniques. If the randomization of X is well-executed, the researcher can derive the causal effect of X on Y simply by comparing the mean value of Y in the two groups. The reason for this is that the random assignment makes the treatment and control groups perfectly comparable (see above). The two are perfectly similar on all relevant variables. Thus, there is no need to control for them in the analysis, as they cannot influence the difference in Y between the two groups. For example, imagine that she finds the average turnout rate

is 60 per cent in the countries of the treatment group and 57 per cent in the countries of the control group. She knows that this difference is due to the electoral system, because that is the only systematic difference between the two groups. She can thus conclude that the average treatment effect is 3 per cent (i.e. 60–57). In other words, if a country changes from a non-proportional to a proportional system, turnout will increase by 3 per cent. Then, she can calculate a simple t-test to evaluate whether this effect is statistically significant or not.

However, experiments also have several limitations. The first, and maybe most important, relates to external validity. Experiments are necessarily somewhat artificial. Firstly, it is the researcher who assigns the variable X to the subjects under study. Experiments, regardless of whether they are in the lab or the field, necessarily deviate from the real-life situation where variable X naturally occurs and where it is not arbitrarily assigned by an external actor. In the example above, it would be artificial to randomly select countries and force them to use one electoral system instead of another one. In reality, the electoral system used in a country is the result of a long political history and a negotiation between key actors at a moment of critical juncture (Bawn 1993).

Secondly, experiments in the lab are even more artificial. As mentioned above, the very goal of lab experiments is to recreate a situation that resembles a real-life one, but in a lab. This recreation necessarily implies a simplification. Even trying very hard, it is impossible for a researcher to recreate a situation that is 100 per cent the same as reality – which often involves a multitude of factors and dimensions. For example, Helios Herrera, Massimo Morelli and Thomas Palfrey (2014) study the effect of electoral systems on turnout using a lab experiment. To do so, they organize elections between subjects in a lab. The subjects vote upon the distribution of a pot of money between them. The main treatment X is whether the election is conducted under a proportional or non-proportional rule. The result is that the turnout rate is higher with a non-proportional system when the voters anticipate that the election is going to be close and higher with a proportional system when this is not the case. The lab experimental design resembles a real election: (1) subjects, just like real-life voters, must cast a vote without knowing what others will do, and (2) the result of this collective vote determines the distribution of a pot of money between them, just as the government

redistributes more or fewer benefits to certain groups of voters. However, the design is also a simplification of a real election where millions of voters cast a vote and where they are also motivated by non-monetary political allegiances. Consequently, the oversimplified design of some experiments can cast doubts on the generalization of findings to explain phenomena that occur outside of the lab. This is the problem of external validity.

That being said, the flip side of the issue of external validity is that, for the reasons mentioned above, experimentalists can be sure that it is the treatment X that changes the outcome Y. They can thus safely interpret the results in terms of causality. In other words, although lab experiments sometimes suffer from a certain lack of external validity, they offer strong guarantees regarding internal validity.³

Another limit of lab experiments is more practical. Lab experiments usually require the researcher to engage in a fastidious process. Before conducting her experiment, she needs to set up a sound research design, which usually implies gathering informed feedback from other experimentalists. If the design is found not to be internally valid, the results might not be meaningful, and it will be too late to change the experiment once it has taken place. The researcher also needs to secure an ethics certificate from her university or lab. Many journals now wish to see the approval of an ethics committee before publishing the results of an experiment.⁴ What is more, the researcher usually needs to secure a grant to cover the participation fee of experimental subjects and the cost of the lab time. It is only at this stage that the experiment can be conducted. However, even at this stage, publication is far from being guaranteed. Sometimes (maybe even often?) experiments lead to null results. For example, treatment X might not have any effect on outcome Y. It is very hard, even impossible, to publish non-results in most scientific journals.

Conducting a lab experiment from start to finish is, thus, a demanding process. However, other empirical observational methods are not always less demanding in terms of time and money. For example, all the steps outlined above also apply to the conduct of an original survey (with the exception that the grant needs to be even larger as surveys tend to be more expensive). Also, qualitative interviews usually require a great investment in time to secure contacts with the targeted population. That being said, interested researchers need to be aware that experiments are no panacea.

ECONOMICS- AND PSYCHOLOGY-BASED EXPERIMENTS

In this section, I argue that most lab experiments in political science can be classified into two ideal-typical groups depending on the scientific discipline that inspired them: those following the tradition of lab experiments in economics⁵ and those following lab experiments in psychology.⁶ The two types are different in terms of: (1) the outcome of interest (behaviour vs. attitude); (2) whether the responses of the subjects to the treatment X are incentivized with money or not; (3) the degree to which the experimental design is an abstraction of the situation studied or a realistic reproduction of it; and (4) whether the focus is on group interactions or individual responses to the treatment X. Table 1 provides a summary of these differences. I go through each line of division, one after the other. I also highlight the respective advantages of the two types of lab experiments. It is important to note that a number of lab experiments do not fall into one of these two ideal-typical categories because they have features of both. I also give examples of these ‘hybrid’ lab experiments below.

The first line of division is the outcome of interest. Experiments of the economics type usually study concrete behaviours as they occur in the lab. For example, Timothy Feddersen, Sean Gailmard and Alvaro Sandron (2009) study whether the probability that a subject will make a difference to the electoral outcome affects the way they vote. To do so, they recruit experimental subjects, put them in a lab, and organize a series of elections between them. The subject’s voting behaviour at these elections constitutes the outcome Y.

In contrast, lab experiments of the psychological type usually focus on reported attitudes. Diana Mutz and Byron Reeves (2005) study how negative and uncivil (that is, ‘dirty’) political debates affect people’s

Table 1
Two Ideal-Types of Lab Experiments

	<i>Economics experiments</i>	<i>Psychological experiments</i>
Outcome of interest	Behaviour	Attitude
Incentives	Monetary	Non-monetary
Design	Abstract	Realistic
Focus	Interactions between subjects	Individual subjects
Examples	Feddersen et al. (2009) Kanthak and Woon (2015)	Mutz and Reeves (2005) Levendusky (2013)

trust in politics. To do so, they ask their subjects, after the experiment, to what extent they agree with statements such as ‘politicians generally have good intentions’ or ‘at present, I feel very critical of the political system’. Then they combine the responses in order to construct an indicator of political trust. This attitude constitutes the outcome Y. Some lab experiments of the psychological type also study behaviour, but in this case they often rely on self-reported behaviour. For example, in their study of name recognition on vote choice, Cindy Kam and Elizabeth Zechmeister (2013) subliminally show the subjects a random name on the screen (treatment X) before asking them whether they would be more willing to vote for a political candidate with this name or another one with a different name (outcome Y). There is no obvious advantage or disadvantage of studying behaviours over reported attitudes (or vice versa). It all depends on the goal of the research.

The second line of division is whether the researcher gives monetary incentives to subjects during the experiment.⁷ At the beginning of most lab experiments, subjects receive a fixed amount of money as a compensation for their time. Typically, in lab experiments of the economics type some subjects then receive more money than others depending on their actions and those of others. The rationale is that subjects reveal their ‘real’ behaviour when there is something at stake (just as in real life, where many actions have real consequences). For example, in the lab elections that Feddersen et al. (2009) organized, subjects could gain money depending on the outcome of the election. Each subject could choose to vote for option A or B. If option B won more votes, everybody received some money but in unequal amounts (some receiving more than others). If option A won more votes, everybody received a little less money, but an equal amount. Their argument is that voting for option A is ‘morally superior’ to voting for option B, but subjects can maximize their self-interest by voting for option B. The treatment X, which the authors randomly assign using a clever design, is the likelihood that each individual’s vote will make a difference to the winning option. The result is that the more unlikely it was to make a difference, the more people voted for the ‘morally superior’ option.

By contrast, in their lab experiment of the psychological type, Mutz and Reeves (2005) do not give any monetary incentives to subjects (other than the fixed compensation at the beginning). They show them a political debate in which actors (realistically) played the role of politicians, and then asked them to complete a questionnaire

about how they felt about politics (see above). The post-experiment questionnaire is in fact a survey questionnaire. However, unlike classic surveys, the respondents watch a visual stimulus – the political debate – before answering the questions. In this experiment, treatment X is the degree of ‘uncivility’ of the political debate presented to the subjects. The result is that watching an uncivil debate diminished people’s trust in politics.

Giving monetary incentives to subjects for their actions has advantages and disadvantages. On the plus side, it allows the researcher to be confident that what she observes in the lab is not just ‘cheap talk’. In the case of Feddersen et al. (2009), if there was no monetary incentive linked to voting, most subjects would have chosen the ‘morally superior’ option in the hope of appearing altruistic to others. On the minus side, depending on the topic, money is not always the most optimal way to incentivize subjects. In elections that occur outside the lab, voters have non-monetary motivations. For example, they are motivated by political allegiances or policy preferences. Mutz and Reeves (2005) did not use any monetary incentive and, hence, had to rely on the honesty of the subjects when they answered the post-experiment questionnaire. However, it is worth noting that researchers make the same assumption about the honesty of respondents when they conduct a traditional survey. This is another way in which the questionnaire used by Mutz and Reeves (2005) is the equivalent of a traditional survey questionnaire.

The third line of division is the distance between the experimental protocol and the concrete situation that the researcher is interested in. In the economics type of lab experiments, the researcher usually tries to construct a design that abstracts from reality. For example, Kristin Kanthak and Jonathan Woon (2015) study why some people choose to become political candidates – that is, why they choose to submit themselves to a selection process in the hope of becoming the representative of a group of people (outcome Y). First, the subjects had to decide whether to volunteer to become a representative of the group, and then they had to select their representative from among the volunteers. Subsequently, the selected representative had to solve some mathematical problems – the more problems she could solve, the more money the other participants received. The experimental design was, thus, an abstraction from the reality of what happens when an individual decides (or not) to become a politician, but it includes some important features of this reality. Politicians are

selected by their fellow citizens and work for others in exchange for some reward. The mathematical tasks represent the effort they make while in office. The result is that when the selection of the representative is decided via an election rather than at random (treatment X), women subjects volunteer less often, even though the election is gender-blind, thus showing that women are more selection-adverse than men.

In contrast, psychological lab experiments usually try to be as close as possible to the real situation they wish to study. For example, Matthew Levendusky (2013) studies how the major media outlets contribute to the polarization of public opinion in the US (outcome Y). To do so, he presented subjects with either recent news capsules from major political TV shows, or very realistic manufactured newspaper editorials (treatment X). He then asked them various questions about their policy preferences. The result was that showing biased news, regardless of the direction of this bias, strongly increased how people felt about a wide range of policies.

Abstract experimental designs have advantages and disadvantages. The key advantage is that the subjects are not influenced by what they think about the world while making their choices in the lab. The goal is to reveal their profoundly human reactions to various situations and to achieve a maximal level of internal validity. In other words, the goal of abstract experiments is to limit the confounding effects of factors that exist outside the lab. For example, in their experiment, Kanthak and Woon (2015) do not provide the opportunity for the group representative to 'make a political career', although this is what the experiment is studying. Consequently, subjects who are interested in politics are no more likely to volunteer than others, and the researchers can reveal more profound motivations (see below). Some lab experiments of the economics type go even further in the level of abstraction. For example, John Duffy and Margit Tavits (2008) sought to evaluate how much voters overestimate their chances of affecting the outcome of an election by organizing elections between subjects in the lab, but by making absolutely no reference to elections in the instructions they give to them.

The disadvantage of abstract experimental designs is the lack of external validity. In the examples presented above, the situation is so abstracted from the reality under study that one might wonder whether the results say anything about this reality. By contrast, experiments of the psychological type seek realism. Consequently,

the researcher can be more confident about the external validity of her study. In his experiment, Levendusky (2013) showed real-life news capsules to the subjects. He could thus be confident that their reaction in the lab was the one they would have had outside of the lab if they had watched these capsules.⁸ However, it is important to note that the external validity is not always guaranteed, even with realistic lab experiments. As James Druckman, Jordan Fein and Thomas Leeper (2012) show, people do not watch just any TV channel; they select the channels that show the news that comforts their existing opinions. To address this issue, they conducted an experiment in which they let the subjects decide which news capsules they wanted to watch, before asking them questions about their policy preference. There is, thus, a trade-off between external and internal validity in both economics and psychological lab experiments.

The final line of division is the focus of the study. The design of lab experiments of the economics type usually involves interactions between subjects. This appears clearly in the design of Kanthak and Woon (2015) that I described above. However, it is important to note that in their study, the outcome *Y* is the decision whether to become a politician or not, which is an individual decision made by each subject separately, even though they make this decision in a context in which they know that the other participants are making similar decisions at the same time. It is not because the experiment involves group interactions that the outcome must be at this level. Many economics lab experiments, however, study outcomes *Y* that are at the level of the group. For example, Damien Bol, André Blais and Simon Labbé St-Vincent (2018) organize elections in the lab in which some subjects play the role of parties and others play the role of voters. The outcome *Y* is the effective number of parties at the elections.

By contrast, experiments of the psychological type are mostly conducted with individual subjects rather than groups.⁹ Here again, there are exceptions. David Sanders (2012) organized deliberative polls between subjects of various European countries, asking them to discuss various topics, such as immigration, and then measuring their policy preferences about these topics. The interactions between subjects are, thus, at the heart of his design. However, despite this particularity, Sanders' (2012) study shares all the features of a psychological lab experiment (attitude as outcome of interest, no monetary incentive, realistic design).

There are, thus, multiple lines of division between lab experiments of the economics and psychological type. A final important remark that I want to make in this section is that most topics can be studied with either type. For example, Richard Lau and David Redlawsk (1997) studied the capacity of voters to identify the ‘correct’ candidate for them – that is, the one that would best serve their interest if elected – using a very typical lab experiment of the psychological type (self-reported behaviours as outcome, no monetary incentives, realistic design and individual focus). Later, André Blais, Simon Labbé St-Vincent, Jean-Benoit Pilet and Rafael Treibich (2016) studied the same topic, using very typical economics experiments (behaviours as outcome, monetary incentives, abstract design and collective focus). It is up to the researcher to decide which type of experiment she wants to use for her research, knowing the advantages and disadvantages of each type of design.

THE CHALLENGES OF LAB EXPERIMENTS

Although lab experiments constitute a powerful tool to estimate the causal effect of a treatment X on an outcome Y, the life of lab experimentalists is not always easy. Lab experiments have attracted some critiques in recent years – critiques that need to be addressed for the method to gain legitimacy. In this section, I discuss three of the most important challenges that lab experimentalists are facing today.¹⁰

A common critique concerns the sample. There is a tendency among lab experimentalists to rely on convenient samples. Often, they recruit university students to participate in their study. It is convenient in the sense that students are already in the campus where the experimental lab is usually situated, and that they only require a small monetary compensation (sometimes they even participate for course credits). The extensive use of student samples in lab experiments has triggered a legitimate external validity critique: would the results of lab experiments be different if they were conducted with a more diverse sample of people (Kam et al. 2007)? To address this critique, some researchers have replicated the same experiments in the lab and in the field (usually with an online survey), using both student and more diverse samples.

On the one hand, the results of classic psychological experiments do not seem to be strongly affected either by the sample type or by the location of the experiment (Clifford and Jerit 2014; Jerit et al. 2013). On the other hand, the results of economics experiments tend to be different when conducted on highly trained economics students, because they are better at (or maybe just more used to) identifying the strategy that maximizes their utility (Belot et al. 2015). However, this difference seems to disappear both when it is very easy to identify the maximizing strategy and when it is very hard to do so, such as in a voting experiment in which subjects simultaneously elect a candidate (Bol et al. 2016). All in all, it seems that the potential bias associated with the use of convenient samples in lab experiments is often overestimated and that observations made on student samples can often be generalized to the rest of the population (Coppock and Green 2015).

A second challenge relates to the ethics of lab experiments. The method raises obvious ethical issues since it involves real human subjects. In assigning treatments to subjects, there is a risk that the researcher increases their level of anxiety, stress or discomfort. It is now widely acknowledged that everything should be done to minimize the negative effects of experiments, making sure that subjects are not from a population at risk (for example, do not suffer from mental health issues), and that they give their full consent before the experiment is conducted.¹¹

The issue of deception – that is, whether the researcher should deliberately lead the subjects to believe something that is not true – has generated a particularly heated debate (McClendon 2012). For example, Ismail White, Chryl Laird and Troy Allen (2014) use a lab experiment to evaluate how much social pressure and self-interest affect the willingness of people to conform to a social norm. During the 2012 US presidential campaign, they recruited African-American students and told them they could choose to distribute a pot of \$100 to either Obama or Romney (the two main presidential candidates). They expected the subjects to favour Obama, because they all supported the Democratic candidate and they knew that their donation was going to be revealed to the other experimental subjects. However, the trick was that for each dollar donated to Romney, the subjects also received a dollar for themselves (and nothing if they chose to donate to Obama). The design allowed the researchers to reveal the tension between self-interest and social pressure but forced

them to use deception. As it is against the law to use public money to make donations to political candidates in the US, they were unable actually to make the donations decided by the subjects. At the end of the experiment, they debriefed them, and acknowledged the deception, justifying it in terms of their research goals.

This example illustrates the multiple problems with deception. First, deception is unethical in itself, as the researchers lie to subjects who consent to participate in their experiments. Second, deception also compromises the work of future lab experimentalists. Imagine a researcher who wants to conduct another experiment with the same pool of subjects as White et al. (2014). These subjects might feel sceptical about this new experiment, knowing that the researcher might be lying to them once again. Their reaction to the treatment will thus be partly affected by whether they suspect the use of deception or not, and the experimental results will be hard to interpret. They might also decline to participate. For these reasons, it is now widely acknowledged that experimentalists in general should, at least as much as possible, not deceive their subjects.

A final critique concerns the reproducibility of lab experiments. The so-called ‘reproducibility crisis’ has hit social sciences, starting with social psychology, and then spreading to other disciplines (Baker 2016). An Open Science Collaboration Project (2015) that involved many researchers throughout the world failed to reproduce a majority of the most-cited and influential lab experiments. It is worrying for social sciences as a whole if what we consider as firmly established results might not be as strong as we thought.

One of the sources of the reproducibility crisis is the research strategy known as ‘fishing expeditions’ or ‘p-hacking’ (Benjamin et al. 2018; Gelman and Loken 2013). This strategy consists of the researcher reporting only a selection of her data and analysis in her paper, to give the impression to the reader that her hypothesis is confirmed and that the treatment effects are statistically significant (hence the expression ‘p-hacking’). It is often driven by the publication bias that many journals hold against experimental results showing the non-effect of the treatment on the outcome. Researchers know that they need to report ‘positive’ findings (effects) if they want to publish their paper and thus engage in fishing expeditions to find some effect.

Some innovations within the political science community aim to address this problem. For example, a community of social science

experimentalists has created the online system Evidence in Government and Politics (EGAP, egap.org), in which researchers can pre-register their experiment.¹² They write an outline detailing their design, including how many subjects they want to recruit and how they intend to analyse the results. The website then saves this document, so that when the paper is submitted for publication in a journal, the reader can easily check whether what is reported fits the initial intentions of the researcher.

It is important to note that the issue of reproducibility is not specific to lab experiments. All studies, including those using observational methods, suffer from problems of reproducibility. Indeed, an advantage of lab experiments is that a researcher can easily try to reproduce them in her own lab. She simply has to implement the same design as the original study. This is not the case for all studies using an observational method. Indeed, studies using a qualitative type are often hardly reproducible at all.

THE FUTURE OF LAB EXPERIMENTS

In order to discuss the future of lab experiments in political science, it is important to look at the past. In the first section of this article, I showed that the number of experimental papers published in the most prominent journals of the discipline has substantially increased within the last 10 years. Other reviews show that the first papers appeared in the 1960s, but the method only really ‘kicked in’ in the late 1990s (Druckman et al. 2011; Morton and Williams 2010).¹³ This is when the experimental turn started. Very few political scientists used experiments in their research before that.

The experimental turn in political science is one realization of a broader trend that has touched all disciplines in social sciences: the ‘credibility revolution’ (Angrist and Pischke 2010). This revolution was driven by a willingness to make political and social research more credible in the eyes of policymakers. Policymakers are primarily interested in how much they can influence the society by their intervention. Hence, they would like to know how much society will change if they implement a certain policy. In other words, they are interested in knowing the causal effect of their potential intervention before making a final decision about it.

As described above, observational methods are not well equipped to identify accurate causal effects. Unless the researcher is able to prove that

there is no issue of reverse causality or omitted variable bias in her analysis, she cannot be certain that what she observes is a true causal estimate. Therefore, policymakers cannot rely on observational studies to make policy decisions. It is too important to rely on conjectural results. This is why experiments are crucial for political scientists: the method is necessary to make political research relevant for the world outside academia. Experiments also make the social sciences more credible in the eyes of other scientists, especially hard scientists, as the experimental method is almost universally considered to be the ideal form of scientific inquiry.

What is the future of lab experiments in political science? The period that directly followed the experimental turn consisted for the most part of replicating existing observational studies with an experimental method. This exercise was necessary to evaluate whether what we thought we knew was genuinely correct and this is certainly one of the reasons for the exponential increase in the number of experimental papers since 1990. So many observational studies needed to be replicated.

For example, experimental studies changed the vision that political scientists had of the relationship between the electoral system and turnout. For a long time, we thought that proportional systems increase turnout. Some observational studies indeed showed evidence pointing in this direction (Blais and Carty 1990). However, recent lab experiments have found that the causal relationship between the electoral system and turnout is more complex. Non-proportional systems actually increase turnout when the election is close (Herrera et al. 2014).

Sometimes experiments confirm the findings of observational studies. Still in the field of electoral systems, a study using observational methods found that electoral systems that combine proportional representation and low district magnitude are ‘sweet spots’, in the sense that they are the best compromise between a fair representation of all the segments of the society in the decision-making process and an efficient accountability mechanism of governments (Carey and Hix 2011). A few years later, a lab experiment replicated this study and found similar results (Labbé St-Vincent et al. 2016), thus confirming the validity of the original observational results.

The initial phase of replication is already well advanced in political science. Lab experimentalists are now exploring new topics that have not been researched before in various fields such as comparative politics and international relations. The method seems to be

particularly prominent in flagship journals of the discipline. Maybe, then, observational methods will become obsolete. It is true that experiments are better equipped to estimate causal effects, which is important for increasing the credibility of the discipline.

However, observational studies are not pointless. Typically, they can be used to overcome the limits of experiments. As mentioned above, experiments are sometimes criticized for their lack of external validity. Lab experiments necessarily imply a simplification of the reality under study, which can cast doubt on the extent to which their results are generalizable to phenomena that occur outside of the lab. A promising way to address this critique is to combine observational and lab experiments in a single paper. For example, Richard Lau and David Redlawsk (1997) use a lab experiment to study ‘correct voting’, understood as a vote for the party/candidate that makes promises that are the closest to one’s beliefs and values. They show that in the lab, a vast majority of voters are able to cast a correct vote, even when they do not have all the information about parties and candidates. To evaluate whether this finding holds in real-life elections, Lau and Redlawsk used survey data from several US presidential elections. They found that the proportion of correct voters is similar to their lab elections. Observational studies can thus be very useful as a complement to lab experiments.

CONCLUSION

In this review, I have examined the use of lab experiments in political science. I have described the advantages and limitations of this method and showed that they can be classified according to two ideal-types: economics and psychology experiments. I further explained the various lines of division between these two types, described the main challenges that political experimentalists are facing today, and briefly discussed the future of experiments in the discipline.

My main argument is that lab experiments are an excellent tool to identify and estimate causal effects. They can be applied to a wide variety of topics and do not require the mobilization of complex statistical techniques. What is more, the well-known problem of external validity of lab experiments can be overcome with better and more diverse samples, and in combination with observational

methods. It is clear that experiments in general and lab experiments in particular have their best days ahead of them.

NOTES

¹ Sometimes the researcher brings the experimental lab to the subjects herself. These ‘lab-in-field’ experiments are usually used when the researcher studies a specific population that is too remote to come to her lab. For example, Gottlieb (2017) uses a lab-in-field experiment to study the impact of local brokers on elections in rural communities in Senegal.

² In this section, I assume that the variable X, in experimental jargon we usually talk about the ‘treatment X’, is binary (presence or absence of X). However, the argument is also valid for categorical variables with more than two categories, or even continuous variables. What matters is that the subjects are randomly assigned to the different categories or values of the treatment X. Similarly, I also assume that the analysis is at the individual level, or ‘subject level’ in experimental jargon. However, the analysis can be at the level of a group of subjects.

³ On the trade-off between external and internal validity, see Schram (2005) or McDermott (2002a).

⁴ On ethics in lab experiments, see below.

⁵ For a specific review of lab experiments of the economics type, see Palfrey (2009).

⁶ It is important to note the ‘economics’ and ‘psychological’ labels are related to the design of the lab experiments, and not their content. Some economics experiments study psychological processes, and some psychological experiments study economics interactions. For example, Duffy and Tavits (2008) use an economics experiment to show that people are not rational when they decide whether to vote in an election, because they are overconfident about their probability of affecting the electoral outcome. This is a psychological process, also called ‘behavioural’ in the economics literature.

⁷ For more details on monetary incentives, see Dickson (2011).

⁸ Here again, the description of the trade-off between external and internal validity is voluntarily exaggerated. It is reasonable to think that the external validity in the study of Levendusky (2013) is not perfect, given that participants are more likely to be attentive when they watch a capsule in a lab experiment than when they watch it in real life.

⁹ See the example of Levendusky (2013) presented above.

¹⁰ For a broader discussion, see McDermott (2002b).

¹¹ For a discussion of the ethics of experiments, see Desposato (2016). Field experiments often raise extra ethical concerns, since the researcher, who directly intervenes in the reality (see above), can affect this reality. For example, there was a scandal in 2014 concerning a field experiment conducted in Montana during the campaign that preceded the election of a new state judge. The researchers sent letters to a random group of voters revealing the ideological position of the

candidates. Some perceived this as an intrusion in the politics of the state, as judges are supposed to remain ideologically neutral (Willis 2014).

¹² Note that, at time of writing this line, EGAP is mostly used to pre-register field experiments.

¹³ There are a few exceptions of political science studies published before 1960, such as Gosnell (1926) or Eldersveld (1956).

REFERENCES

- Adida, C.L., Laitin, D.D. and Valfort, M.-A. (2016), ‘“One Muslim is Enough!” Evidence from a Field Experiment in France’, *Annals of Economics and Statistics*, 121/122: 121–60.
- Amorim Neto, O. and Cox, G.W. (1997), ‘Electoral Institutions, Cleavage Structures, and the Number of Parties’, *American Journal of Political Science*, 41(1): 149–74.
- Angrist, J.D. and Pischke, J.-S. (2010), ‘The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics’, *Journal of Economic Perspectives*, 24(2): 3–30.
- Baker, M. (2016), ‘1,500 Scientists Lift the Lid on Reproducibility’, *Nature*, 533(7604): 452–4.
- Bawn, K. (1993), ‘The Logic of Institutional Preferences: German Electoral Law as a Social Choice Outcome’, *American Journal of Political Science*, 37(4): 965–89.
- Belot, M., Duch, R. and Miller, L. (2015), ‘A Comprehensive Comparison of Students and Non-students in Classic Experimental Games’, *Journal of Economic Behavior and Organisation*, 113(1): 26–33.
- Benjamin, D.J., Berger, J.O., Johannesson, M. et al. (2018), ‘Redefining Statistical Significance’, *Nature Human Behavior*, 2: 6–10.
- Blais, A. and Carty, K.R. (1990), ‘Does Proportional Representation Foster Voter Turnout?’, *European Journal of Political Research*, 18(2): 167–81.
- Blais, A., Labbé St-Vincent, S., Pilet, J.-B. and Treibich, R. (2016), ‘Voting Correctly in Lab Elections with Monetary Incentives: The Impact of District Magnitude’, *Party Politics*, 22(4): 544–51.
- Boix, C. (1999), ‘Setting the Rules of the Game: The Choice of Electoral Systems in Advanced Democracies’, *American Political Science Review*, 93(3): 609–24.
- Bol, D. (2019), ‘Experiments: A Tool to Test Causal Relationships’, in F. Morin, C. Olsson and E. Ozlem Atkcan (eds), *Key Concepts in Research Methods* (Cambridge: Routledge).
- Bol, D., Labbé St-Vincent, S. and Lavoie, J.-M. (2016), ‘Recruiting for Laboratory Voting Experiments: Exploring the (Potential) Sampling Bias’, in A. Blais, J.-F. Laslier and K. Van der Straeten (eds), *Voting Experiments* (New York: Springer): 271–86.
- Bol, D., Blais, A. and Labbé St-Vincent, S. (2018), ‘Which Matters Most: Party Strategic Exit or Voter Strategic Voting? A Laboratory Experiment’, *Political Science Research and Methods*, 6(2): 229–44.
- Carey, J.M. and Hix, S. (2011), ‘The Electoral Sweet Spot: Low Magnitude Proportional Electoral Systems’, *American Journal of Political Science*, 55(2): 383–97.

- Clifford, S. and Jerit, J. (2014), 'Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies', *Journal of Experimental Political Science*, 1(2): 120–31.
- Coppock, A. and Green, D.P. (2015), 'Assessing the Correspondence between Experimental Results Obtained in the Lab and Field: A Review of Recent Social Science Research', *Political Science Research and Methods*, 3(1): 113–31.
- Desposato, S. (2016) (ed.), *Ethics in Experiments: Problems and Solutions for Social Scientists and Policy Professionals* (London: Routledge).
- Dickson, E.S. (2011), 'Economics Versus Psychology Experiments: Stylization, Incentives, and Deception', in J.N. Druckman, D.P. Green, J.H. Kuklinski and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science* (Cambridge: Cambridge University Press): 58–71.
- Dickson, E.S., Gordon, S.C. and Huber, G.A. (2015), 'Institutional Sources of Legitimate Authority: An Experimental Investigation', *American Journal of Political Science*, 59(1): 109–27.
- Druckman, J.N., Green, D.P., Kuklinski, J.H. and Lupia, A. (2011), 'Experimentations in Political Science', in J.N. Druckman, D.P. Green, J.H. Kuklinski and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science* (Cambridge: Cambridge University Press): 3–13.
- Druckman, J.N., Fein, J. and Leeper, T.J. (2012), 'Source of Bias in Public Opinion Stability', *American Political Science Review*, 106(2): 430–54.
- Duffy, J. and Tavits, M. (2008), 'Beliefs and Voting Decisions: A Test of the Pivotal Voter Model', *American Journal of Political Science*, 52(3): 603–18.
- Eldersveld, S.J. (1956), 'Experimental Propaganda Techniques and Voting Behavior', *American Political Science Review*, 50(1): 154–65.
- Feddersen, T., Gailmard, S. and Sandron, A. (2009), 'Moral Bias in Large Elections: Theory and Experimental Evidence', *American Political Science Review*, 103(2): 175–92.
- Gartner, S.S. (2008), 'The Multiple Effects of Casualties on Public Support for War: An Experimental Approach', *American Political Science Review*, 102(1): 95–106.
- Gilligan, M.J., Pasquale, B.J. and Samii, C. (2014), 'Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal', *American Journal of Political Science*, 58(3): 604–19.
- Gelman, A. and Loken, E. (2013), 'The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "P-Hacking" And the Research Hypothesis Was Posited Ahead of Time', mimeo, Columbia University.
- Gosnell, H.F. (1926), 'An Experiment in the Stimulation of Voting', *American Political Science Review*, 20(4): 869–74.
- Gottlieb, J. (2017), 'Explaining Variation in Broker Strategies: A Lab-in-the-Field Experiment in Senegal', *Comparative Political Studies*, 50(11): 1556–92.
- Herrera, H., Morelli, M. and Palfrey, T. (2014), 'Turnout and Power Sharing', *Economic Journal*, 124(574): 131–62.
- Jerit, J., Barabas, J. and Clifford, S. (2013), 'Comparing Contemporaneous Laboratory and Field Experiments on Media Effects', *Public Opinion Quarterly*, 77(1): 256–82.
- Kam, C.D. and Zechmeister, E.J. (2013), 'Name Recognition and Candidate Support', *American Journal of Political Science*, 57(4): 971–86.

- Kam, C.D., Wilking, J.R. and Zechmeister, E. (2007), "Beyond the Narrow Data Base": Another Convenience Sample for Experimental Research', *Political Behavior*, 29(4): 415–40.
- Kanthak, K. and Woon, J. (2015), 'Women Don't Run? Election Aversion and Candidate Entry', *American Journal of Political Science*, 59(3): 595–612.
- King, G., Keohane, R.O. and Verba, S. (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press).
- Labbé St-Vincent, S., Blais, A. and Pilet, J.-B. (2016), 'The Electoral Sweet Spot in the Lab', *Journal of Experimental Political Science*, 3(1): 75–83.
- Lau, R.R. and Redlawsk, D.P. (1997), 'Voting Correctly', *American Political Science Review*, 91(3): 585–98.
- Levendusky, M.S. (2013), 'Why Do Partisan Media Polarize Viewers?', *American Journal of Political Science*, 57(3): 611–23.
- Lijphart, A. (1971), 'Comparative Politics and the Comparative Methods', *American Political Science Review*, 65(3): 682–93.
- Lowell, A.L. (1910), 'The Physiology of Politics', *American Political Science Review*, 4(1): 1–15.
- McCauley, J.F. (2014), 'The Political Mobilization of Ethnic and Religious Identities in Africa', *American Political Science Review*, 108(4): 801–16.
- McClelland, G. (2012), 'Ethics of Using Public Officials as Field Experiment Subjects', *Newsletter of the APSA Experiments Section*, 3(1): 13–20.
- McDermott, R. (2002a), 'Experimental Methodology in Political Science', *Political Analysis*, 10(4): 325–42.
- McDermott, R. (2002b), 'Experimental Methods in Political Science', *Annual Review of Political Science*, 5: 31–61.
- Morton, R.B. and Williams, K.C. (2010), *From Nature to the Lab: The Methodology of Experimental Political Science and the Study of Causality* (Cambridge: 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(1): 1–15.
- Open Science Collaboration (2015), 'Estimating the Reproducibility of Psychological Science', *Science*, 349(6251): aac4716.
- Palfrey, T.R. (2009), 'Laboratory Experiments in Political Economy', *Annual Review of Political Science*, 12: 379–88.
- Przeworski, A. and Teune, H. (1970), *The Logic of Comparative Social Inquiry* (Oxford: Wiley Interscience).
- Ragin, C.C. (1989), *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press).
- Sanders, D. (2012), 'The Effects of Deliberative Polling in an EU-wide Experiment: Five Mechanisms in Search of an Explanation', *British Journal of Political Science*, 42(3): 617–40.
- Sauermann, J. and Kaiser, A. (2010), 'Taking Others into Account: Self-interest and Fairness in Majority Decision Making', *American Journal of Political Science*, 54(3): 667–85.
- Schram, A. (2005), 'Artificiality: The Tension Between Internal and External Validity in Economic Experiments', *Journal of Economic Methodology*, 12(2): 225–37.

- Terris, L.G. and Tykocinsky, O.E. (2016), 'Inaction Inertia in International Negotiations: The Consequences of Missed Opportunities', *British Journal of Political Science*, 46(3): 701–17.
- White, I., Laird, C.N. and Allen, T.D. (2014), 'Selling Out? The Politics of Navigating Conflicts between Racial Group Interest and Self-interest', *American Political Science Review*, 108(4): 783–800.
- Willis, D. (2014), 'Professors' Research Project Stirs Political Outrage in Montana', *New York Times*, 29 October.