EXPERIMENTS

Damien Bol, King's College London

EXPERIMENTAL AND OBSERVATIONAL METHODS

For decades, social scientists were convinced that experimentations were not for them. Although they generally acknowledged the great merits of the method (see **POSITIVISM**), they thought they could not use it, for practical and ethical reasons. Where it is easy for natural scientists to manipulate non-living elements such as metals, social scientists considered that they could not and should not experiment with their objects of study. Consequently, the use of **COMPARATIVE ANALYSIS** was recommended as a substitute. The comparison of different-yet-similar people or countries was perceived as the best way to test **HYPOTHESES**, and in particular those positing a causal relationship between a **VARIABLE** X and an outcome Y. This idea is at the root of many social sciences methods such as qualitative **CASE STUDIES** or quantitative **REGRESSION ANALYSIS**, which I will call '**OBSERVATIONAL METHODS**' hereafter.

Yet, since 1990, experiments have become increasingly popular in the social sciences. In experiments, the researcher randomly assigns a **VARIABLE** X to some of the cases, but not all of them. She can thus observe the value of the outcome Y in the cases that received X and compare it to the other cases that did not. This contrasts with **OBSERVATIONAL METHODS**, for which the researcher simply observes the variations of X and Y as they occur in reality.

Experiments have two important advantages compared to **OBSERVATIONAL METHODS**. First, they allow the researcher to clearly identify what the causal **VARIABLE** X is and the outcome Y. This distinction is made by the researcher prior to the analysis, which discards the possibility of reverse causality (see **CAUSATION**). Second, with **OBSERVATIONAL METHODS** the precision of the estimates depends on the extent to which the researcher manages to control for the differences between the cases. When she cannot entirely capture these differences, the estimates are likely to be inflated, underestimated, or simply wrong.

EXPERIMENTS AND THE POTENTIAL-OUTCOME FRAMEWORK

Most social sciences researchers aim at testing the causal relationship between X and Y (see **CAUSATION**). They seek to estimate how much Y increases/decreases when X changes. Here I am considering that X is a binary variable (i.e. an event that either happens or not), and Y is a continuous variable. However, the arguments are also valid for other types of variables.

Traditionally, we approach causal relationships via the 'Neyman-Rubin potential-outcome framework' (Rubin 1979). In this framework, if a researcher wants to

estimate the causal effect of X on Y, she needs to observe, for the same case, the value of Y when X happens, and the value of Y when X does not happen. Then, by subtracting one value from the other, she has a precise estimate of how much the value of Y changes when X happens. However, she cannot observe the two outcomes at the same time. In reality, either X happened or not; a given person cannot experience both X and non-X.

Let me give an example: imagine you are interested in testing the causal relationship between going to university (X) and happiness (Y). You observe the level of happiness of a group and know who among them went to university. You might be tempted to estimate the causal effect by comparing the level of happiness of those who went to university to that of those who did not. However, this would not be a meaningful estimate as there are many other **VARIABLES** that influence the level of happiness. Some of these variables directly relate to the probability of going to university, such as whether people grew up in a rich family. Unless you have the capacity to control for all these variables, you do not know whether it is going to university, or any of these other variables, that causes happiness.

What the potential-outcome framework says is that you need to observe, within the group of people who went to university, what would have been their level of happiness if they did not go (and vice versa for those who did not go to university). This is obviously impossible. Yet you can approach the ideal potential-outcome framework by randomizing the **VARIABLE** X (which then becomes 'treatment X'). Typically, this entails randomly splitting the cases into two groups, and deciding that for one of these groups X happens, whereas for the other group, it does not. Then a simple comparison of the average value of Y in the two groups will give the average treatment effect: for all the cases under study, if X happens, Y changes, on average, by the average treatment effect.

Let me come back to the example: imagine you are capable of randomly splitting the group into two and assigning the treatment X 'going to university' to only one of them, making sure that the others do not go. You can then measure the average level of happiness in the two groups and subtract them. This group comparison becomes a meaningful estimate of the causal effect thanks to the random assignment of the treatment X. There are two reasons for that. First, it removes the possibility that other **VARIABLES** that influence X in real life have a disturbing effect (e.g. having rich parents). Second, if the groups are sufficiently large, the differences in Y that are due to variables other than the treatment will cancel each other out (see **STATISTICAL SIGNIFICANCE**).

The way I have presented experiments here is similar to the way they are presented in most textbooks. Yet this is a narrow definition of the method. Sometimes, textbooks present a stretched definition that includes 'natural' and 'quasi-experiments'. The term natural experiments refers to experimental situations that naturally occur in reality. The treatment is randomly assigned by someone other than the researcher. An example of natural experiments is the 1969 draft lottery that was used to determine which young American men would be called to fight in Vietnam. In their paper, Erikson and Stoker (2011) compare the individuals who were randomly selected by the lottery to those who were not (treatment X). They find, among other things, that participation in the Vietnam War caused people to be, years later, more anti-war (outcome Y).

The term quasi-experiments refers to situations in which the treatment X is not completely randomly assigned, but in which this assignment is 'as if' it was random.

An example of quasi-experiments is the study of Hainmueller, Hangartner, and Pietrantuono (2015), which shows that immigrants tend to be more politically active (outcome Y) when they obtain the nationality of the country in which they live (treatment X). In some cantons in Switzerland, naturalization is decided by popular referendums. The authors compare levels of political activity of immigrants who received Swiss citizenship by just a few votes over the majority threshold at these referendums, to those who did not by just a few votes below the threshold. Since only a few votes separate the two, the assignment of the treatment X was as if random (maybe this could have been simply a mistake in counting votes). Natural and quasi-experiments are not experiments *stricto sensu*, as it is not the researcher who directly assigns the treatment to individuals.

EXPERIMENTS IN THE FIELD AND IN THE LAB

There are two broad types of experiments: experiments in the field (including survey experiments), and in the lab. In field experiments, the researcher goes out in the field and randomly assigns the treatment X to some individuals as they are living their normal life. Subsequently, she observes their behaviour, and how they respond to the treatment. By contrast, in lab experiments, the researcher recruits some participants and brings them into a lab. It is in this lab that she randomly assigns the treatment X and observes the reaction of the subjects.

Let me give examples of both field and lab experiments. Gerber, Green, and Larimer (2008) conducted a field experiment in Michigan shortly before the 2006 US primary election. They identified about 150,000 people who were registered as voters. Then, they randomly split these potential voters into several groups, and assigned them different treatments X, which consisted of various letters encouraging them to vote. Observing the turnout rate in each of the experimental groups (outcome Y), they found that social pressure causes an increase in turnout rate. Turnout is on average 8 percentage points higher among people who received a letter reminding them that in the US everybody can see who did or did not vote.

In another study, Adida, Laitin, and Valfort (2016) used a lab experiment to test whether religious difference causes prejudice. They recruited French and Senegalese participants and brought them into a lab in Paris. They divided the participants into random groups and asked them whether they were willing to give up to 5€ to other people in their group, or to keep this money for themselves. They found that French participants donated less money when there were many Senegalese participants in their group, especially if these Senegalese participants had a Muslim first name. The average treatment effect of adding a second Muslim into the group was a decrease in donation of 1.23€. They concluded that religious difference in groups causes discriminatory behaviours.

Both field and lab experiments have advantages. Field experiments are more realistic in the sense that they build on real social behaviours. Their results thus have more external validity compared to lab experiments, in which social interactions are abstracted

via lab interactions. For example, Gerber, Green, and Larimer (2008) observed actual turnout rates which are their study topic, whereas Adida, Laitin, and Valfort (2016) studied donation into a lab as a proxy for prejudice against Muslim people. However, in lab experiments, the researcher has more control over her experimental design, and especially the assignment of the treatment. Gerber, Green, and Larimer (2008) did not know whether the participants opened the letter they sent them, whereas Adida, Laitin, and Valfort (2016) showed the names and faces of each participant in the group to make sure that this piece of information was known to all. The choice of field and lab experiments is a trade-off between external validity and the researcher's control.

TOWARDS ETHICAL EXPERIMENTS

Experiments are the ideal tools to test causal relationships between a **VARIABLE** X and an outcome Y. However, they also have limits. Experiments cannot be used to study all social sciences topics. There are some **VARIABLES** X that the researcher cannot randomly assign to individuals. Coming back to the example regarding the effect of going to university on happiness, it seems impossible for a researcher to impose that some people go to university and others not. This impossibility is even clearer when it comes to macro phenomena like revolutions, wars, or famine. Yet, this does not mean that the researcher cannot experimentally study these topics. She can, for example, conduct a lab experiment in an area with a recent history of civil war, to evaluate the level of social cohesion among individuals (Gilligan, Pasquale, and Samii 2014).

The question of which topic can be analysed with experiments revolves around the question of **ETHICS IN RESEARCH**. It is crucial that the researcher interested in conducting an experiment thinks through all potential consequences the research could have on the people involved. It is about making sure that it will not create too much harm, anxiety, stress, or embarrassment.

REFERENCES

Adida, Claire L., David D. Laitin, and Marie-Anne Valfort. 2016. 'One Muslim is Enough! Evidence from a Field Experiment in France'. *Annals of Economics and Statistics* 121-122: 121–160.

Erikson, Robert S. and Laura Stoker. 2011. 'Caught in the Draft: The Effects of Vietnam Draft Lottery Status on Political Attitudes'. *American Political Science Review* 105(2): 221–237.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. 'Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment'. *American Political Science Review* 102(1): 33–48.

Gilligan, Michael J., Benjamin J. Pasquale, and Cyrus Samii. 2014. 'Civil War and Social Cohesion: Lab–in–the–field Evidence from Nepal'. *American Journal of Political Science* 58(3): 604–619.

Hainmueller, Jens, Dominik Hangartner, and Giuseppe Pietrantuono. 2015. 'Naturalization Fosters the Long-term Political Integration of Immigrants'. *Proceedings of the National Academy of Sciences* 112(41): 12651–12656.

Rubin, Donald B. 1979. 'Using Multivariate Sampling and Regression Adjustment to Control Bias in Observational Studies'. *Journal of the American Statistical Association* 74(366): 318–328.