Experimental Political Science and the Study of Causality

From Nature to the Lab

Rebecca B. Morton

Kenneth C. Williams

The Causal Inference Problem and the Rubin Causal Model

In this book we concentrate on two prominent approaches to causality that underpin almost all of the work in political science estimating causal relationships: the Rubin Causal Model (RCM) and the Formal Theory Approach (FTA).[1] In this and the next two chapters, we focus on the RCM model. In Chapter 6 we discuss FTA. RCM is an approach that has its genesis in the statistical literature and early studies of field experiments,[2] whereas FTA comes primarily from the econometric literature and is the approach used by experimental economists in many laboratory experiments (although statisticians such as Pearl [2000] are also advocates of an approach that is similar to FTA).

We begin our presentation of RCM by defining some basic variables and terms used in the approach.

---

[1] There are other approaches to causality; see Winship and Morgan (1999) for a review of the literature and Dawid (2000) for a critique of current approaches. RCM and FTA are by far the most prevalent ones used in political science.

[2] Although RCM is named after his work, Rubin (1990) notes that the ideas behind the approach precede his work and were first applied to experimental research (see Neyman, 1923). Because of Neyman's contribution, some call the model the Neyman–Rubin Causal Model. Rubin receives due credit for providing a formal foundation for use of the approach with observational data. However, the use of counterfactuals and potential outcomes can also be found in formal models of the early econometrics literature (see Roy, 1951; Quandt, 1958, 1972; and Lewis, 1963). Heckman (2005a) reviews this literature in economics and also points to early work in mathematical pyschology on counterfactuals (Thurstone, 1930). For a review of the statistical literature, see Rubin (1990), Winship and Morgan (1999), and Wooldridge (2002, chapter 18). Sometimes the model is also called the Rubin–Holland approach in recognition of Holland's noteworthy (1986) formalization. (Holland is credited with naming the model after Rubin.)

3.1 Variables in Modeling the Effects of a Cause

3.1.1 The Treatment Variable

We are interested in the effect of information on how voters choose in the election studied or target population; that is, information is our possible causal variable. To simplify our analysis, we think of being informed as a binary variable (later we discuss analyses where this assumption is relaxed). We can think of there being two states of the world, one in which an individual i is informed about the choices before them in the election and the other in which the individual i is uninformed. In a naturally occurring election, the information the voter may or may not have may be about the policy positions of the candidates on the issues they care about or some overall quality of the candidates that matters to voters such as honesty, integrity, or the ability to manage the economy or national defense. In the referenda example, the voter may or may not have information on aspects of the proposed law. And in the laboratory election, similarly, the subject may or may not have information about the choices before him or her in the election.

In general, we can denote the two states of the world that a voter can be in as "1" or "0" where 1 refers to being informed and 0 refers to being uninformed or less informed. Let $T_i = 1$ if an individual is in state "1"; $T_i = 0$ otherwise. So in an experiment like Example 2.4, in which Mondak et al. provide subjects with information about hypothetical candidates' qualities, 1 would mean that a subject read and understood the information provided and 0 would mean otherwise.

Typically we think of T_i as the treatment variable. In political science we often refer to it as the main or principal independent variable. We are interested in the effect of the treatment variable on voting choices.

Definition 3.1 (Treatment Variable): *The principal variable that we expect to have a causal impact.*

3.1.2 Variables That Affect Treatments

Because we are interested in both experimental and observational data, it is useful to think of T_i as an endogenous variable, which is a function of a set of observable variables, Z_i; a set of unobservable variables, V_i; as well as sometimes a manipulated variable, M_i, which may be manipulated by a researcher or by nature. The manipulated variable describes whether the

subject was assigned to view a campaign advertisement by the researchers. In a laboratory experiment, as in that of Dasgupta and Williams, the manipulated variable is whether subjects were told the quality of the candidates directly by the experimenter. We call this an experimental manipulation. In observational data without manipulation, T_i is only a function of Z_i and V_i; we call this a natural manipulation.

Definition 3.2 (Manipulated Variable): *A variable that has an impact on the treatment variable that can be manipulated either naturally or by an experimenter.*

Definition 3.3 (Experimental Manipulation): *When the manipulated variable is fixed through the intervention of an experimentalist.*

Definition 3.4 (Natural Manipulation): *When the manipulated variable is part of nature without intervention of an experimentalist.*

The observable variables, Z_i, are observable confounding factors as defined in Definition 2.8, and the unobservable variables, V_i, are unobservable confounding factors as defined in Definition 2.9. In the case of an experiment during an election in the field, as in Clinton and Lapinski's experiment (see Example 2.3), observable variables might be the individuals' level of education, their income, or their place of residence whereas unobservable variables might be the individuals' cognitive abilities, interest in politics, or free time.[3]

3.1.3 The Dependent Variable

We are interested in the effects of a cause and we call the dependent variable the variable that represents or measures the effects, essentially. In our running example of information and voting, the dependent variable is the voting behavior upon which we expect information to have an effect. Whether informed or not, our individuals have choices over whether to vote or abstain, and if they vote, which candidate or choice to vote for. If our target election is a U.S. presidential election with three candidates, then the voter has four choices: abstain, vote for the Republican candidate, vote for the Democratic candidate, or vote for the minor party or independent

---

[3] Of course we may attempt to measure some of these unobservables and make them observable; the point is that there are always some influences that will be unobservable.

candidate. So in standard political science terminology, our dependent variable is a random variable, Y_i, that can take on four values $\{0, 1, 2, 3\}$, where 0 denotes individual i choosing abstention, 1 denotes individual i voting for the Republican, 2 denotes individual i voting for the Democrat, and 3 denotes the individual voting for the minor party or independent candidate. Denote Y_{i1} as the voting choice of i when informed and Y_{i0} as the voting choice of i when uninformed.

Definition 3.5 (Dependent Variable): *A variable that represents the effects that we wish to explain. In the case of political behavior, the dependent variable represents the political behavior that the treatment may influence.*

Note that in some of the experiments in our examples in the appendix to the Chapter 2, the dependent variable is a survey response by subjects as to how they would choose rather than their actual choice, since observing their actual choice was not possible for the researchers.

We hypothesize that Y_{ij} is also a function of a set of observed variables, X_i, and a set of unobservable variables, U_i, as well as T_i. For example, Y_{ij} might be a function of a voter's partisan affiliation, an observable variable, and it might be a function of a voter's value for performing citizen duty, an arguably unobservable variable. Note that we assume that it is possible that Z_i and X_i overlap and could be the same and that U_i and V_i overlap and could be the same (although this raises problems with identification of causal relationships, as discussed in the following chapters).

3.1.4 Other Variables and Variable Summary

Occasionally we also discuss two other variables: W_i, the set of all observable variables, and P_i, the probability that $Y_i = 1$. We define and use W_i when the analysis makes no distinction between X and Z. In some of the examples and sometimes for ease of exposition, we focus on the case where Y_i is binary – it can be either 0 or 1. This might be the case where the dependent variable is turnout and 0 represents abstention and 1 represents voting as in the study of Lassen (2005) discussed in the next chapter. In other situations, 0 might represent voting for a Democratic candidate and 1 voting for the Republican, as in the analysis of Bartels (1996) discussed in the next chapter. When the choice is binary, the dependent variable is often assumed to be the probability that Y_i equals 1, which we designate as P_i. Denote P_{i1} as the value of P_i when i is informed and P_{i0} as the value of P_i when i is uninformed. Table 3.1 presents a summary of this notation, which is used throughout Chapters 3–5.

Table 3.1. *Variable definitions*

Variable	Definition
T_i	Treatment of unit of observation i In our example, $T_i = 0$ if an individual is uninformed, $T_i = 1$ if an individual is informed
Z_i	Observable variables that partly determine T_i
V_i	Unobservable variables that partly determine T_i
M_i	Manipulated variable that partly determines T_i This could be manipulated by an experimenter or nature
Y_i	The actual outcome or choice of the unit In our example, the vote choice of the individual
Y_{ij}	The hypothetical outcome or choice of the unit when $T_i = j$ In our example, the hypothetical vote choice of the individual when $T_i = j$ (which may not be observed)
X_i	Observable variables that partly determine Y_{ij} There may be overlap between X_i and Z_i
U_i	Unobservable variables that partly determine Y_{ij} There may be overlap between V_i and U_i
W_i	$X_i \cup Z_i$
P_i	The probability that $Y_i = 1$ when $Y_i = \{0, 1\}$
P_{ij}	The hypothetical value of P_i when $T_i = j$

3.2 Manipulations Versus Treatments

3.2.1 Why Are Manipulations Sometimes Called Treatments?

Many would argue that in the ideal experimental manipulation $M_i = T_i$, and Z_i and V_i have no effect on T_i. But this ideal experimental manipulation is impossible when we think about voter information as the treatment and, as we argue later, may not be the best option for the study under question. But even if we have this as a goal, that our manipulation variable is equivalent to our treatment variable, it is not possible when examining human behavior. The most ideal case for such a situation would be in the laboratory, but even there we find disconnects between the two. For instance, in Example 2.6, the subjects' education and income levels, cognitive abilities, interest in the laboratory election, and belief about the experimenter's truthfulness can vary and affect whether the subjects are able to comprehend the information provided to them by the experimenter. In general, for experiments in social sciences, there are usually some observable and unobservable aspects of humans that affect our ability to manipulate the treatments the subjects experience.

Often experimentalists, ourselves included, use the term treatment to refer to the manipulated variable itself or a particular experimental manipulation. For example, Gerber, Kaplan, and Bergan in Example 2.1 refer to the groups to which subjects are assigned as "treatment" groups. When an experimentalist does so, then the implicit presumption is that the manipulated variable is the treatment and that the researcher has complete control over the treatment variable. Because we know that the treatment and manipulation are two different things – that voter information is affected by observable and unobservable variables independent of the manipulation in the experiment – why would an experimentalist ignore this reality? The answer lies in random assignment of the manipulations. In Chapter 5 we explore in more detail how well random assignment works.

3.2.2   When Treatment Variables Cannot Be Manipulated

When we think of treatment variables, we also allow for those that arguably cannot be manipulated through experimentation (or can only be manipulated with great difficulty). Thus, it is possible in our model of causality for the treatment to be things like gender, race, ethnicity, and so on. This is in contrast to Holland (1986, 1988), who argues that there can be no causal effect of gender on earnings because it is not possible to assign gender to subjects. However, it is possible to assign subjects with particular genders to different work situations and measure their earnings. Consider, for example, the experiment of Michelitch (2010) in Ghana. She investigates the effects of ethnicity on bargaining behavior. Specifically, when two individuals are bargaining over a price, say a taxi ride, if they are in the same ethnic group, is the price lower? In her experiment she assigns people to bargaining situations by their ethnic type to determine the answer to her question. Although she cannot assign ethnicity to an individual, she can choose the combinations of ethnic identities of the individuals in the bargaining situations.

Heckman (2005a,b) and Heckman and Vytlacil (2007a,b) point out, and we agree, that the view that ethnicity cannot be a treatment variable conflates the difficulty in identifying causal effects with defining causal effects. It may be that identifying the causal effect of gender, race, or ethnicity on something like earnings, in political science voting, is difficult, but that does not mean it is not an interesting question that we can imagine asking.

3.2.3   When Manipulations Affect Treatments Only Indirectly

In the example experiments we have discussed so far, the manipulations all affect the treatment variable in a straightforward manner. In Example 2.6,

the manipulation directly provides subjects with information about the true jar; in Example 2.1, the manipulation directly provides subjects with newspapers containing information about the election; in Example 2.3, the manipulation directly provides subjects with campaign advertisements; and in Example 2.8, the manipulation directly provides subjects with information through the experience of living with the policy experiment.

In some experiments, the researcher engages in manipulations, but the manipulations do not directly affect the treatment variable the researcher is focusing on in a straightforward way. Example 3.1 presents two studies where the researchers use an experimental laboratory to investigate voter brain responses to candidate appearances using functional magnetic resonance imaging (fMRI) equipment. The treatment variables that are under study by the researchers (candidate appearances) are not directly manipulated.

Example 3.1 (Brain and Candidate Image Lab Experiment): *Spezio et al. (2008) report on an experiment using brain imaging (fMRI) equipment to determine whether positive or negative attributions of candidate images play a primary role in how candidate appearance affects voting.*

Target Population and Sample: Spezio et al. did not report how their subjects were recruited. They used two separate samples of participants in California. In study 1 they used 24 subjects aged 18–38, seven of which were female. In study 2 they used 22 white women aged 20–35, who were all registered to vote and had voted in one or more of the following national elections: 2000, 2002, and/or 2004. In both studies the participants had no history of neurological or psychiatric illness and were not on antipsychotic medications. The participants also had no prior knowledge of any of the political candidates whose images were used and reported no recognition of the candidates.

Compensation: Spezio et al. did not report if the subjects received compensation for their participation.

Environment: The experiments were conducted at the California Institute of Technology using a Siemens 3.0-T Trio MRI scanner.

Procedures: The researchers conducted two studies. In study 1, subjects were shown "200 grayscale images of political candidates who ran in the real 2006 U.S. midterm elections for either the Senate (60 images), the House of Representatives (74 images), or Governor (66 images). The stimuli were collected from the candidates' campaign Web sites and other Internet sources. An electoral pair consisted of two images of candidates, one Republican and one Democrat, who ran against one another in the real election. Due to the racial and gender composition of the candidates, 70 of the 100 pairs were of male politicians, and 88 of 100 pairs involved two Caucasian politicians. An

independent observer classified 92% of the images as 'smiling'. In 57% of the pairs, both candidates were frontal facing, in the rest at least one was facing to the side. Except for transforming color images into a gray scale, the stimuli were not modified. Images were presented using video goggles. . . .

The study was conducted in the month before the 2006 election. An effort was made to avoid pairs in which one of the candidates (e.g. Hillary Clinton) had national prominence or participated in a California election, and familiarity ratings collected from all of the participants after the scanning task verified the stimuli were unfamiliar. . . .

Participants were instructed that they would be asked to vote for real political candidates who were running against each other in the upcoming midterm election. In particular, they were asked to decide who they would be more likely to vote for given that the only information that they had about the politicians were their portraits.

Each trial consisted of three events. . . . First, a picture of one of the candidates was centrally presented for 1 s. Second, after a blank screen of length 1–10 s (uniform distribution), the picture of the other candidate in the pair was presented for 1 s. Third, after another blank screen of length 1–10 s, the pictures of both candidates were presented side by side. At this point, participants were asked to cast their vote by pressing either the left or right button. They had a maximum of 2 s to make a decision. Participants made a response within this time frame in 100% of the trials. Trials were separated by a 1–10 s blank screen. The order of presentation of the candidates as well as their position in the final screen was fully randomized between participants" (pp. 350–351).

Similarly, study 2 used "60 grayscale images of smiling political candidates who ran in real U.S. elections for the House of Representatives or Senate in either 2000, 2002 or 2004 (30 pairs of opponents)." The images were a subset of those used in a previous study of candidate images on voter choices by Todorov et al. (2005) for comparative purposes. The images were selected such that *"both images in an electoral pair (i) were frontal facing, (ii) were of the same gender and ethnicity and (iii) had clear, approximately central presentation of faces that were of approximately the same size."* Again the pairs matched Republicans and Democrats who had actually run against each other. *"Due to the racial/ethnic and gender composition of the original image library, all stimuli were of Caucasian politicians, and 8 of the 30 pairs were of female politicians. Stimuli were preprocessed to normalize overall image intensity while maintaining good image quality, across all 60 images. All images were presented centrally, via an LCD projector and a rear-projection screen, onto a mirror attached to the MRI head coil, approximately 10 inches from a participant's eyes. . . . A pilot behavioral study confirmed that the social*

judgments made about our selected stimuli were representative of the entire set of face stimuli from which they were drawn. . . .

Participants were instructed that they would be asked to make judgments about real political candidates who ran against one another in real elections. They were told that they would only be given the images of the politicians to inform their judgments. Image order was counterbalanced across participants. Participants made judgments about candidates' attractiveness (Attr), competence (Comp), public deceitfulness (Dect) and personal threat (Thrt) in four separate scanning sessions" (pp. 352–353). Specifically, the participants were asked which candidate in a pair looked more physically attractive to them, more competent to hold national office, more likely to lie to voters, and more likely to act in a physically threatening manner toward them. *"Each session took approximately 9 min to complete"* (p. 353).

Spezio et al. used a protocol that had been used successfully in prior studies of fact preference. That is, "[e]ach trial in a decision block consisted of the sequential presentation of two images in an electoral pair, image A then image B, until a participant entered a decision about the pair via a button press. . . . An A/B cycle on a given trial proceeded as follows: (i) central presentation of a fixation rectangle that surrounded the area in which an image was to appear; (ii) after 4–6 s, a 30 ms display of image A surrounded by the fixation box, accompanied by a small black dot in the lower left corner (indicating that this was image A); and (iii) after 3–4 s, a 30 ms display of image B surrounded by the fixation box, accompanied by a small black dot in the lower right corner (indicating that this was image B). Cycles were separated by 4–6 s and continued until a participant entered a button press or until 30 s had elapsed, whichever came first (no participant ever took the 30 s). Participants were asked to attend overtly to the space inside the rectangle in preparation for a candidate image." The authors used eye tracking to ensure that participants were looking at the stimuli.

Results: In study 1, Spezio et al. found that images of losing candidates elicited greater brain activation than images of winning candidates, which they contend suggests that negative attributions from appearance exert greater influence on voting than do positive. In study 2, Spezio et al. found that, when negative attribution processing was enhanced under the threat judgment, images of losing candidates again elicited greater brain activity. They argue that the results show that negative attributions *"play a critical role in mediating the effects of appearance on voter decisions, an effect that may be of special importance when other information is absent."*

Comments: In study 2, the researchers had to reject the neuroimaging data from 6 participants due to excessive motion. The behavioral data of these

participants were not significantly different from the 16 used in the analysis, however.

Is Example 3.1 an experiment? Certainly it does not fit what some would consider a classic experiment because the treatments investigated – candidate images – are not manipulated directly by the experimenters. The authors have subjects experience a large number of choices and make multiple judgments in study 2, but they do not manipulate those choices to investigate their hypotheses and instead measure the correlation between brain activity and votes and judgments made by the subjects.

Yet we consider it an experiment because the researchers intervene in the data generating process (DGP) and exert control over the choices before the subjects, as discussed in Section 2.4.2. There is no perfect or true experiment. The appropriate experimental design depends on the research question, as is the case with observational data. In fact, the variety of possible experimental designs and manipulations is in some ways greater than the range of possibilities with observational data, as we discuss. It is true, as we show in the following chapters, that when a researcher is investigating the effects of a particular cause, then having the manipulation directly affect the treatment variable (i.e., the proposed causal variable) provides advantages to the researcher in identifying that causal relationship. And it is true that Spezia et al. lost those advantages by not directly manipulating the treatment variable in this fashion.

3.3 The Rubin Causal Model

We have now defined the terms of our model of the effects of a cause. Usually the next step is to discuss how to measure the causal effect of treatments on the dependent variable (i.e., the effect of T_i on Y_i, within a given target population). But before moving on to the issues of identification and estimation, we address the theoretical foundations of studies of causality in political science – the foundations behind different identification and estimation strategies. It is important to discuss these foundations because of the underlying assumptions and the implications these assumptions have for the interpretation of estimated results.

3.3.1 Defining Causal Effects

Our first step is to formally define what we are interested in, the causal effect. In RCM, the causal effect of the treatment for each individual is defined