

PS 0700

Research Designs in Political Science: Experiments

Political Science Research Methods

Professor Steven Finkel

Fall Semester 2022

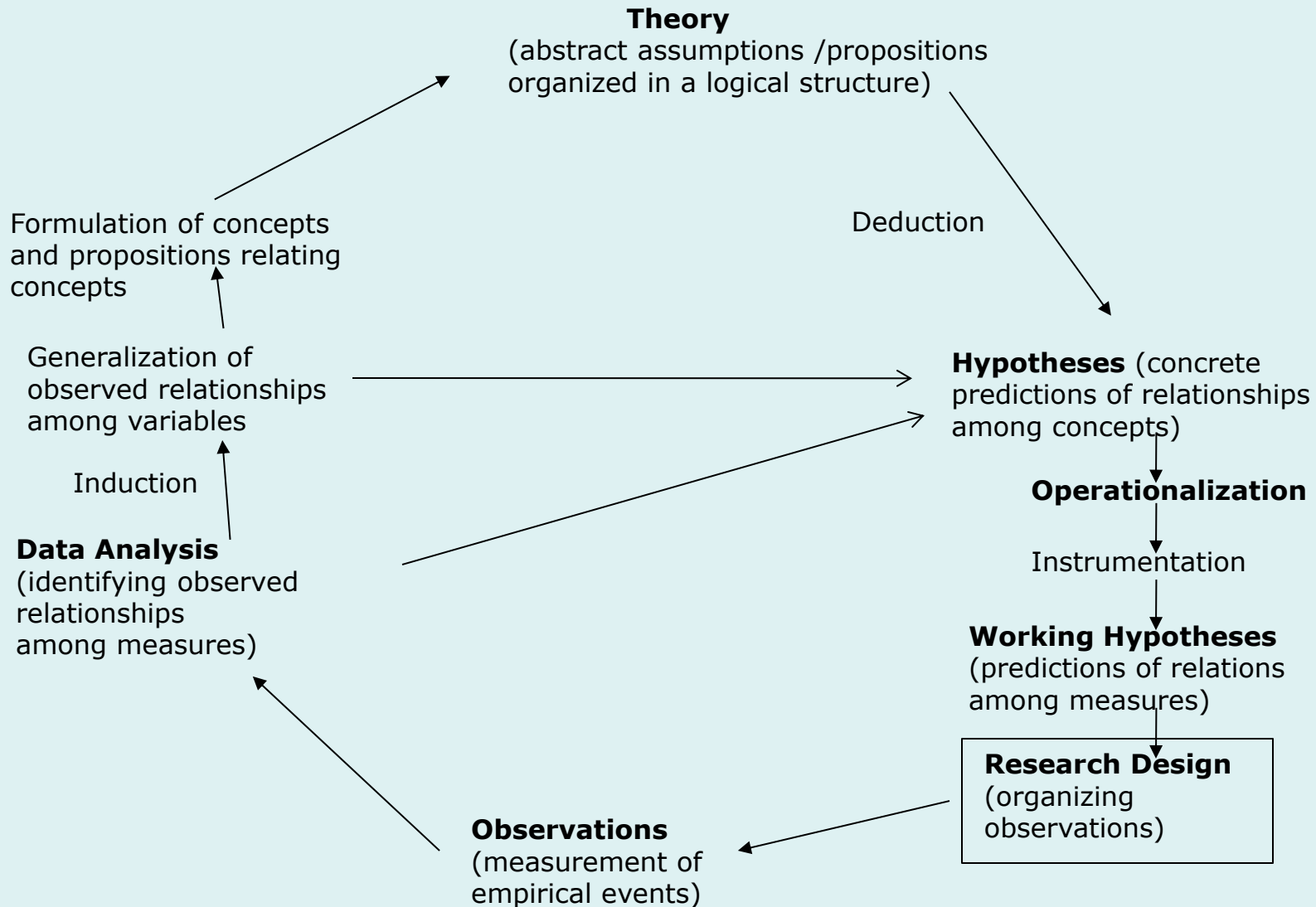
Week 4



Goals for the Session

- Discuss goals of research design
- Discuss experimentation as the “ideal” design for establishing causal relationships
- Discuss drawbacks of experiments in generalizing causal effects, and problems with experiments in “real-world” political research

A Model of the Research Process



“Research Designs involve setting up a research project so that research questions can be answered as unambiguously as possible. The objective of a good research design is to establish causal relationships and to assess their generalizability.”

Meier, Brudney, and Bohte, *Applied Statistics for Public and Non-Profit Administration*, p.52.

- Quotation invokes two critical ideas in setting up a successful research design:
 - Will the study be able to establish a causal relationship between the independent and dependent variables in the context of the specific research setting? If so, the design will have what is called *“internal validity”*
 - Can the results be generalized to other cases or units, and to other settings? If so, the design will have what is called *“external validity”*
 - IMPORTANT NOTE: “internal validity” of a research design is *not* the same as the “validity” of an indicator in terms of measurement
- The ideal design would maximize both internal and external validity. But these designs are incredibly rare in political science (and elsewhere!!!)

The Classic Experimental Design

		Pre-Test			Post-Test	Difference
Treatment Group	R	M_{1t}	X		M_{2t}	$M_{2t} - M_{1t}$
Control Group	R	M_{1c}			M_{2c}	$M_{2c} - M_{1c}$

Where:

M stands for “measurement of the dependent variable Y”;

X stands for “stimulus, or exposure to the independent variable”;

R stands for “randomization”

In short: 1. Create two groups through random assignment; 2) measure the dependent variable; 3) expose one group to the independent variable; and 4) measure the dependent variable again

Critical Features of the Classic Experiment

- The groups are *randomly assigned*
- The researcher controls the treatment, i.e., the form and level of X that is introduced to the “treatment group”
- The groups are measured *before* (“pre-test”) and *after* (“post-test”) the introduction of X to only one of the groups
- The “causal effect” of the treatment is calculated as the “difference in differences” between the two groups, i.e.
$$(M_{2t} - M_{1t}) - (M_{2c} - M_{1c})$$

What is so great about experiments?

INTERNAL VALIDITY: EXPERIMENTS ARE THE STRONGEST DESIGN -- BY FAR -- IN ESTABLISHING CAUSALITY IN THE SPECIFIC CONTEXT

- **Randomization *ensures* non-spuriousness.** In the long run, i.e., with a large enough number of subjects, randomization ensures that the treatment and control group are *statistically equal* before the introduction of X to the treatment group. There are no possible confounding Zs, given appropriate randomization!
- Randomization controls for possible confounding from Zs that we have measured and potentially *could* include in the analysis (if we thought to do so), **as well as** Zs that we have not measured, not observed, or have not even thought of as possible confounders!!
- This means experimental designs control for possible confounding due to the “*observables*” as well as the “*unobservables*”

Experiments and “Counterfactual” Causal Logic

- Experimental designs illuminate the modern “counterfactual” logic of causality in the social sciences in even deeper ways
- Remember the definition of a causal effect from the counterfactual perspective: comparing the *same unit’s (potential) level* of some outcome (Y) under **both treatment and control** conditions. The difference in the (potential) outcome Y that the unit would obtain under “treatment” or “no treatment” is the *causal effect of the treatment* for that individual unit.
- **BUT: the causal effect is unobservable for any given unit**, since a unit either **DOES** or **DOES NOT** receive the treatment, not both!!!
- That is what we have called **the “Fundamental Problem of Causal Inference”**
- We want to estimate causal effects, but they are intrinsically unobservable!!
- How do experiments help with this problem?

- It would be ideal if the groups that we *do* observe, i.e., those who receive and those who do not receive treatment, could serve as appropriate *counterfactuals* or proxies for one another
- As noted in the last lecture, this is *only* possible when all potential Z variables which may produce differences between the groups before we introduce treatment are controlled, or accounted for!!!
- **In experimental designs, this is satisfied through the randomization process!** The two groups – treatment and control – are **exactly** similar (in large enough samples), and **therefore the control group serves as an exact counterfactual proxy for what the treatment group *would have looked like in the absence of treatment*.** Therefore the difference-in-differences between treatment and control produces the causal effect of X
- As we will see, we will need to make additional (sometimes untestable) assumptions or take additional statistical steps to assert that control groups are good counterfactual proxies to treatment groups in *non-experimental* or *observational* designs – but in those kinds of designs it is always possible we will fall short!

What's so great about experiments? (continued)

- Having a **randomly assigned control group** almost certainly overcomes several very serious threats to internal validity and successful causal inference:
 - **History:** something might happen between “pre-test” and the “post-test” aside from X, and this “something” could then possibly cause Y, not X. But since the randomly-assigned control group would also experience this “something”, the classical experiment handles this potential problem!
 - **Maturation:** time itself, not X, may cause individuals to change on Y. But since the randomly-assigned control group would also change (presumably at the same rate given randomization), the classical experiment handles this potential problem!
 - **Reactivity to Pre-Test:** the pre-test measurement may affect the post-test measurement, so that being tested initially on some political measure may affect their responses at a later time. But since the randomly-assigned control group would also be exposed to the pre-test, the classical experiment handles this potential problem!

What is so great about experiments? (continued)

- Since we measure the groups *after* we administer X to one of them, we know that the independent variable X comes *before* we record the changes in the dependent variable Y for both groups. **Therefore we have established “time precedence” between X and Y!**
- So: if we find an effect of X on changes in Y in the classic experimental design, **it *must have been* X that produced it!!!!** Thus this design satisfies **every** criterion for a *causal* effect that we have discussed **and** fulfills the causal requirements from the counterfactual framework as well
- Finally, experiments have a straightforward logic, so it is relatively easy to present the results, especially to policy-makers (i.e. no complicated statistics needed!)

Examples

- Classic Laboratory-Based Experiments
 - Studies of News Media Effects (Iyengar and Kinder, *News that Matters*)
 - Studies of Impact of Campaign Advertising (e.g., Ansolabahere and Iyengar classic study *Going Negative*)
- Field Experiments: Basic or Theoretical Research
 - Voter Turnout and Effects of Mobilization (Green and Gerber, *APSR* 2000)
 - Former Pitt Professor Victoria Shineman on Mobilization and Voter Information (*British Journal of Political Science* 2016)
 - Inoculation Against Misinformation (Roozenbeek *et al.*, *Science Advances* 2022)
- Field Experiments: Policy Research
 - Tennessee Class Size Experiment (Project Star) from 1980s
 - Poverty Action Laboratory (PAL at MIT) – see “Random Acts” on Web site (for recitation discussion)
- Survey Experiments (e.g. Kuklinski *et al.* 1997, see slide 17)

Variations of the Classic Experimental Design

- Possible problem in the classic design: the pre-test “interacts” with X, the experimental stimulus. For example, people who were given a pre-test on their vote intentions might be especially likely to be affected by a stimulus such as a campaign advertisement, and that *without the pre-test*, the effect of the exposure to the advertising would not show up. So the pre-test can lead to a confounding causal effect aside from the “true” effect of X
- Two Solutions: 1) The Solomon Four Group Design, and 2) The “Post-Test Only” Design

The Solomon 4-Group Design

			Pre-Test		Post-Test
Treatment Group 1	R		M _{1t}	X	M _{2t}
Treatment Group 2	R			X	M _{3t}
Control Group 1	R		M _{1c}		M _{2c}
Control Group 2	R				M _{4c}

- Two treatment, two control groups: one of each gets a pre-test
- Compare difference between treatment group 1 over time and control group 1 over time, and then compare to difference between treatment group 2 and control group 2. This is the difference between the groups when there is a pre-test and when there is not. If they differ, it is due to the pre-test interacting with X in treatment group 1

The Post-Test Only Design

		Post-Test	
Treatment Group	R	X	M_t
Control Group	R		M_c

- Also eliminates pre-test bias by excluding it altogether
- Much more common in political science than Solomon 4 Group, mainly due to fewer costs and complexity
- Used often in media studies and in “survey experiments”
- Problem: With no pre-test, we are not **100%** sure that the groups were not different on Y already, *before* the introduction of X.
Randomization *should* work but due to random chance in sampling it may not work perfectly

Last thing that is so great about experiments: excellent unobtrusive data on sensitive topics!

Now I am going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. I don't want to know which ones, just HOW MANY.

With these ground rules established, the interviewer then reads a list of three items:

- (1) the federal government increasing the tax on gasoline;
- (2) professional athletes getting million-dollar contracts;
- (3) large corporations polluting the environment.³

Some randomly assigned respondents receive the baseline version. Others, in the test condition, receive the three baseline items plus a fourth, in this case “a black family moving in next door.”

TABLE 1
ESTIMATED MEAN LEVEL OF ANGER OVER A BLACK FAMILY MOVING IN
NEXT DOOR, BY REGION (WHITES ONLY)

Condition	Non-South	South
Baseline	2.28 (.77) <i>n</i> = 425	1.95 (.80) <i>n</i> = 139
Black family	2.24 (.95) <i>n</i> = 461	2.37 (1.0) <i>n</i> = 136
Estimated % angry	0	42***

Note: Entries are means; standard deviations are in parentheses.

****p* < .001 for regional difference in estimated percent angry (one-tailed test).

Source:
Kuklinski *et al.* (1997),
“Racial
Attitudes
and the
New
South”,
*Journal of
Politics*

- Gonzalez Ocantos, *et al.*, “Vote Buying and Social Desirability Bias: Evidence from Nicaragua”, *AJPS* (2012)

I’m going to hand you a card that mentions various activities, and I would like for you to tell me if they were carried out by candidates or activists during the last electoral campaign. Please, do not tell me which ones, only **HOW MANY**.

For the control group, the following campaign activities are listed and read to respondents:

- they put up campaign posters or signs in your neighborhood/city;
- they visited your home;
- they placed campaign advertisements on television or radio;
- they threatened you to vote for them.

The treatment group is shown and read a fifth category, placed in the third response position:

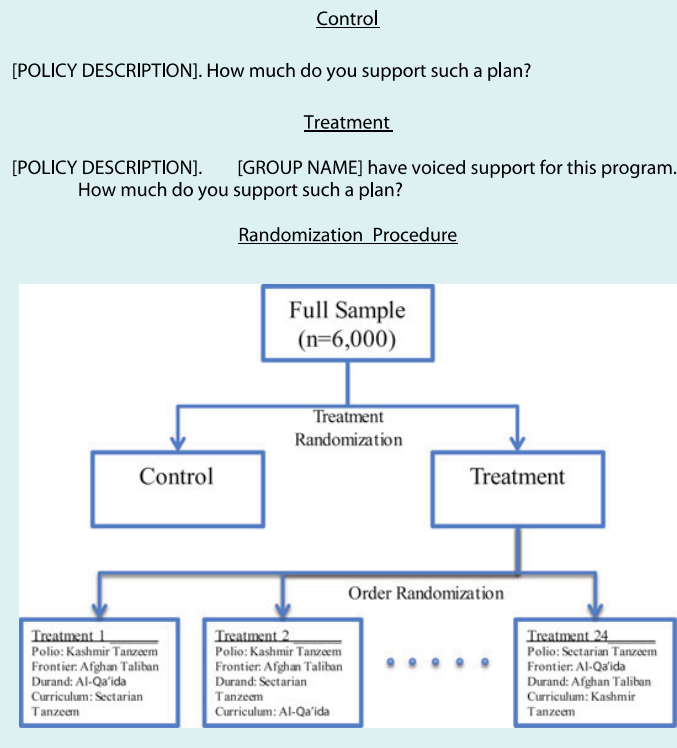
- they gave you a gift or did you a favor⁷

Screensho

- Result: 24% received a gift in exchange for a vote according to the List Experiment, while 2% admitted this via self-report

- Blair *et al.*, “Poverty and Support for Militant Politics: Evidence from Pakistan”, *AJPS* (2012)

FIGURE 1 Illustration of the Endorsement Experiment



- Result: 10-20% less support for Al-Qaeda and Afghan Taliban than in self-reported surveys

Potential Problems with Experiments

- Experiment May Not Detect a Causal Effect Because:
 - Time lag for X to affect Y may not match the design
 - People may change behavior as a result of knowing or suspecting that they are being treated (or not being treated)
 - Outside the controlled setting of the laboratory, there may be problems in implementation:
 - Differential “attrition” (drop-out) of subjects
 - “Contamination” of the treatment and control groups through “spillover” effects (e.g., treated individuals influencing controls)
 - Imperfect compliance with experimental manipulation (some in the experimental group may not “take up” the treatment, and some in the control group may). This is not fatal but requires additional analyses and limits the generalizability of the effects

Biggest Potential Problem for Experiments: Generalizability or “External Validity”

- Many experiments use college students as subjects – will the effects hold for different populations? Maybe not, especially in political applications
- Are the experimental manipulations similar to individuals’ experience with the treatment in the “real world”? Maybe not
- Is the level/amount of the independent variable used in the laboratory the same as the level/amount needed to exert an impact in the “real world”? Alternatively, would implementing the same level/amount of the treatment in the real world produce the same effect as seen in the experiment? Maybe not, which could introduce “scale bias”
- In field experiments, would the results look the same if the experiment were to be conducted in another country, another context? Or are the effects limited to specific subjects in specific time and place?

Other Issues with Experiments

- Many (most?) important explanatory factors in political science are not easy to manipulate – and some are nearly **impossible** to manipulate
 - We cannot assign countries, e.g., to experience terrorism or to have different institutions (but maybe we can simulate this in the lab?)
 - We cannot *assign* individuals, e.g., to get richer/poorer (but maybe lotteries can proxy!)
- There may be serious ethical issues involved in withholding or giving treatments
- Experiments can be costly, esp. outside the lab
- In policy research, need to ask:
 - Are the effects of changing or implementing a new policy or program “worth it” in terms of costs and benefits?
 - *Why* did the treatment work? (May need qualitative data to answer this too)
- Replication of experimental results is often difficult to achieve

Two Really Bad Designs (Why?)

- Post-Test Only, No Control Group

X **M₂**

- Rival explanations: “X did not cause Y because...”

Two Really Bad Designs (Why?)

- Post-Test Only, No Control Group

X **M₂**

- Rival explanations: “X did not cause Y because...”
 - The group was already high (low) on Y *before* X was introduced
 - The group would have registered a high (low) level of Y *regardless* of whether or not X was introduced
 - The group was affected by some Z that took place around the time of X, not by X itself
 - The group tended to be high or low on some other variable Z, and Z was responsible for the level of Y, not X

- Pre-Test-Post-Test, No Control Group

M_1 X M_2

- Rival explanations: “X did not cause Y because...”

- Pre-Test-Post-Test, No Control Group

M_1 X M_2

- Rival explanations: “X did not cause Y because...”
 - The group would have registered a high (low) change in Y *regardless* of whether or not X was introduced
 - The group was affected by some Z that took place around the time of X, not by X itself
 - The group tended to be high or low on some other variable Z, and Z was responsible for the change in the level of Y, not X
 - The pre-test measurement itself led to the change in Y, not the introduction of X