

Social Science Inquiry II

Week 1: Course introduction, part II

Molly Offer-Westort

Department of Political Science,
University of Chicago

Winter 2023

Is social science *science*?

- ▶ In discussing methods for social science research, we should think about what social science is, and what its goals are.
- ▶ What is science?
- ▶ What is social science?
- ▶ Is social science *science*? Why?

Definition of scientific research [KKV]

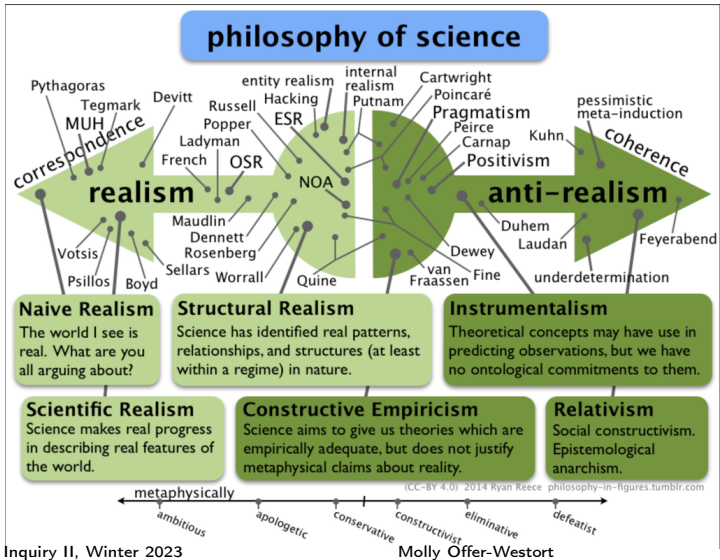
1. The goal is inference.
2. The procedures are public.
3. The conclusions are uncertain.
4. **The content is the method.**

Definition of social science [KKV]

“Social science constitutes an attempt to make sense of social situations that we perceive as more or less complex.”

Philosophy of science

- ▶ These terms do not have definitions that are universally agreed on.
- ▶ ...and there is a whole field on the philosophy of science.



Approach

- ▶ We will focus on study that is empirical, where scientific knowledge is built based on observed, measured phenomena.
- ▶ There are other approaches.

The content is the method.

- ▶ Claim: Research is science if it uses a scientific method.

Path of research development

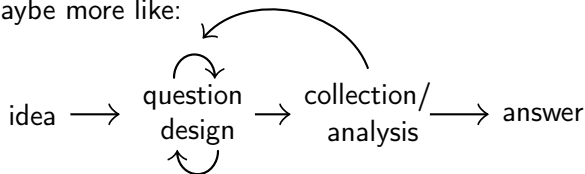
A common linear paradigm for a research project:

question → design → collection → analysis → answer

Tukey (1980) framing:



Maybe more like:



Proposing theories [KKV]

- ▶ Propose theories that are *falsifiable*.
- ▶ If you want to be able to test this theory, it should have *observable implications*. (The more the better.)
- ▶ Clarity and precision in your theory is helpful.

How does this inform how we think about data?

- ▶ Data is not something that is handed down to us in its complete and final form—we select and shape our data.

Principles of data collection [KKV]

1. Record and report data generation process.
2. Collect data on as many observable implications as possible.
3. Maximize the validity of measurements.
4. Ensure that data-collection methods are reliable.
5. Data and analyses should be as replicable as possible.

Observable implications

- ▶ What are these?
- ▶ How do observable implications relate to causal inference?

Inference

- ▶ What is descriptive inference?
- ▶ What is causal inference?

Elements of (descriptive/causal) inference [Holland]

- ▶ Define a population of interest. U
- ▶ Determine a variable that is defined over this population, for which there is variation. A
- ▶ Define response variable “of interest.” Y

We can consider how A and Y vary together. Without further assumptions, this is just correlation.

What does it mean for something to cause something else?

What does it mean for something to cause something else? [Rubin]

- ▶ Possibility for multiple values of treatment—for *each individual*. Treatment *could have been different*. $S(u) = t$ OR $S(u) = c$
- ▶ Some temporal element: outcome variable is realized after “exposure” to treatment variable. $S(u) \rightarrow Y(u)$
- ▶ Conceptual possibility for multiple different versions of response variable—for *each individual*. $Y_t(u)$ may not equal $Y_c(u)$

Consider the individual *causal effect*

$$Y_t(u) - Y_c(u).$$

No causation without manipulation [Holland]

- ▶ Causes are only things that could (hypothetically) be treatments in experiments.
- ▶ NOT attributes—e.g., race and biological sex can't plausibly be manipulated, because there is no meaningful *counterfactual*.
 - ▶ What we can manipulate: others' *perception* of race/sex (Bertrand and Mullainathan, 2004)
 - ▶ To the extent that sex/gender are social constructs, can we consider counterfactual socialization?

Fundamental problem of causal inference

- ▶ We only see the response variable under one version of treatment.
- ▶ Why does this matter?

Resolution (?) of the fundamental problem of causal inference

- ▶ Causal inference is not impossible.
- ▶ But making causal inferences without making assumptions IS impossible.
- ▶ So we need to depend on some assumptions.
- ▶ The crux: how plausible are these assumptions?

Some special cases of causal inference

- ▶ Temporal stability
- ▶ Causal transience
- ▶ Unit homogeneity
- ▶ Constant causal effect
- ▶ Independence

Temporal stability and causal transience

- ▶ The value of response will not change based on *when* you apply treatment to an observation.
- ▶ The value of response will not change if you had at some point previously applied a different treatment to an observation.

Unit homogeneity

- ▶ The value of response under a given treatment is the same for two observations; and the value of response for a different treatment is also the same.
- ▶ You can observe response on the comparable units under one of each version of treatment.

Constant causal effect

- ▶ The difference between the value of the response variable under one version of treatment as compared to another version of treatment is the same for *every* observation.
- ▶ How does this relate to the unit homogeneity assumption?
- ▶ Does this allow us to back out the value of the causal effect, without other assumptions?

Independence

- ▶ There are a “large” number of observations.
- ▶ On average, the units that are assigned one version of treatment look like the units that receive another version of treatment.
- ▶ (What does $E[Y_s|S = t]$ mean?)

The special role of experiments

- ▶ In experiments, treatment is randomly assigned by the researcher.
- ▶ So we know independence holds by design.
- ▶ We can compare groups and get the *average* causal effect.
- ▶ (When can we also get the *individual* causal effect?)

Group discussion

- ▶ Is social science *science*? Justify your answer.
- ▶ What does it mean for something to cause something else? Come up with your own definition.
- ▶ Holland emphasizes focusing on the effects of causes, rather than the causes of effects. Why do you think this is? Do you think there are cases when we should focus on the latter? If so, when?
- ▶ Give an example when you think each of the below causal assumptions would hold.
 - ▶ Temporal stability
 - ▶ Causal transience
 - ▶ Unit homogeneity
 - ▶ Constant causal effect
 - ▶ Independence

References I

- Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. American economic review, 94(4):991–1013.
- Tukey, J. W. (1980). We need both exploratory and confirmatory. The American Statistician, 34(1):23–25.