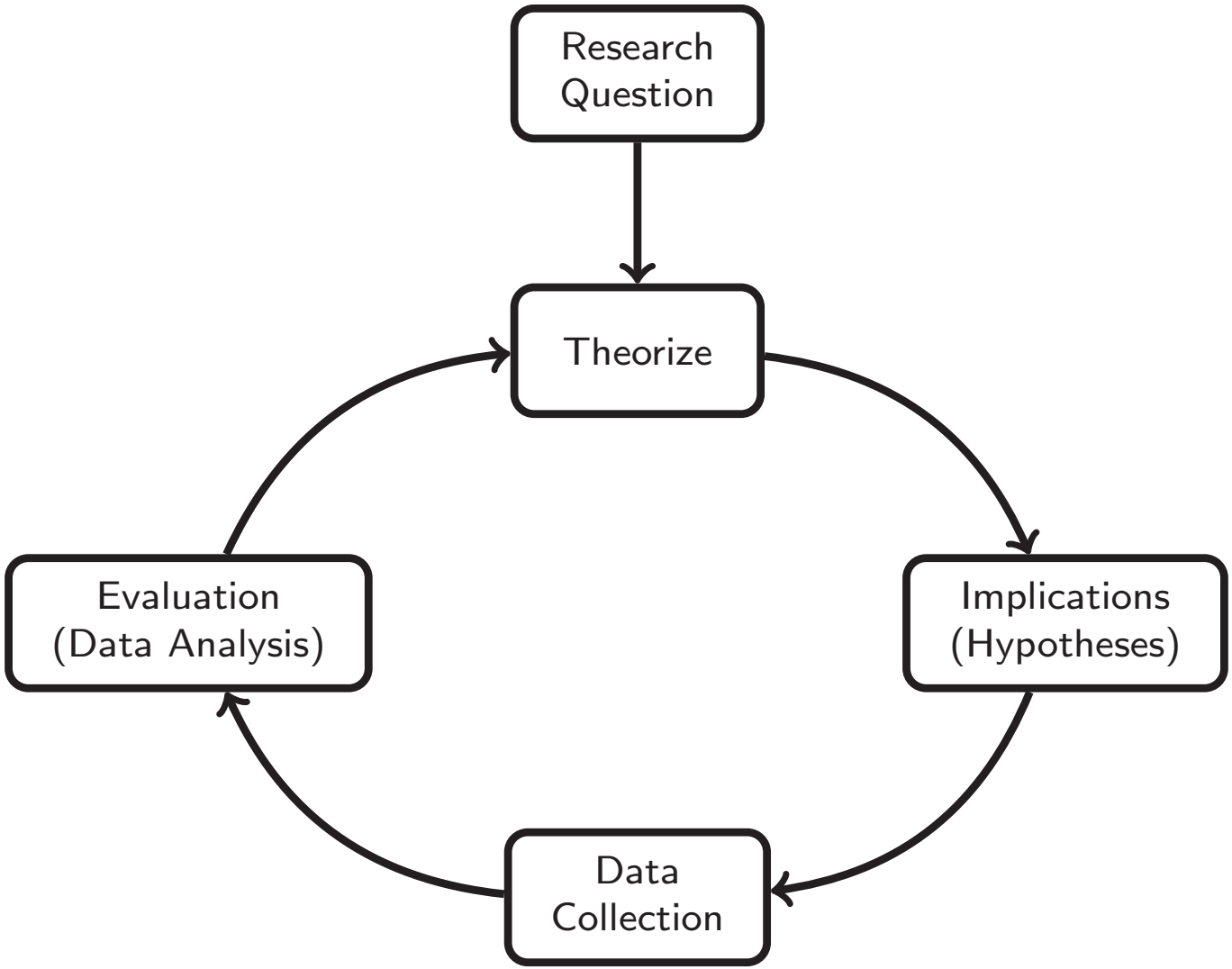
# Week 2: Elements of the Research Process

POLS0007
Principles of Social Science Research

University College London

1. Repetition
   - Research question
   - Theory & model
   - Operationalization and hypotheses
   - Data collection and evaluation
2. The Four Hurdles
3. 5th hurdle: Measurement
4. Conclusion

# The Scientific Process



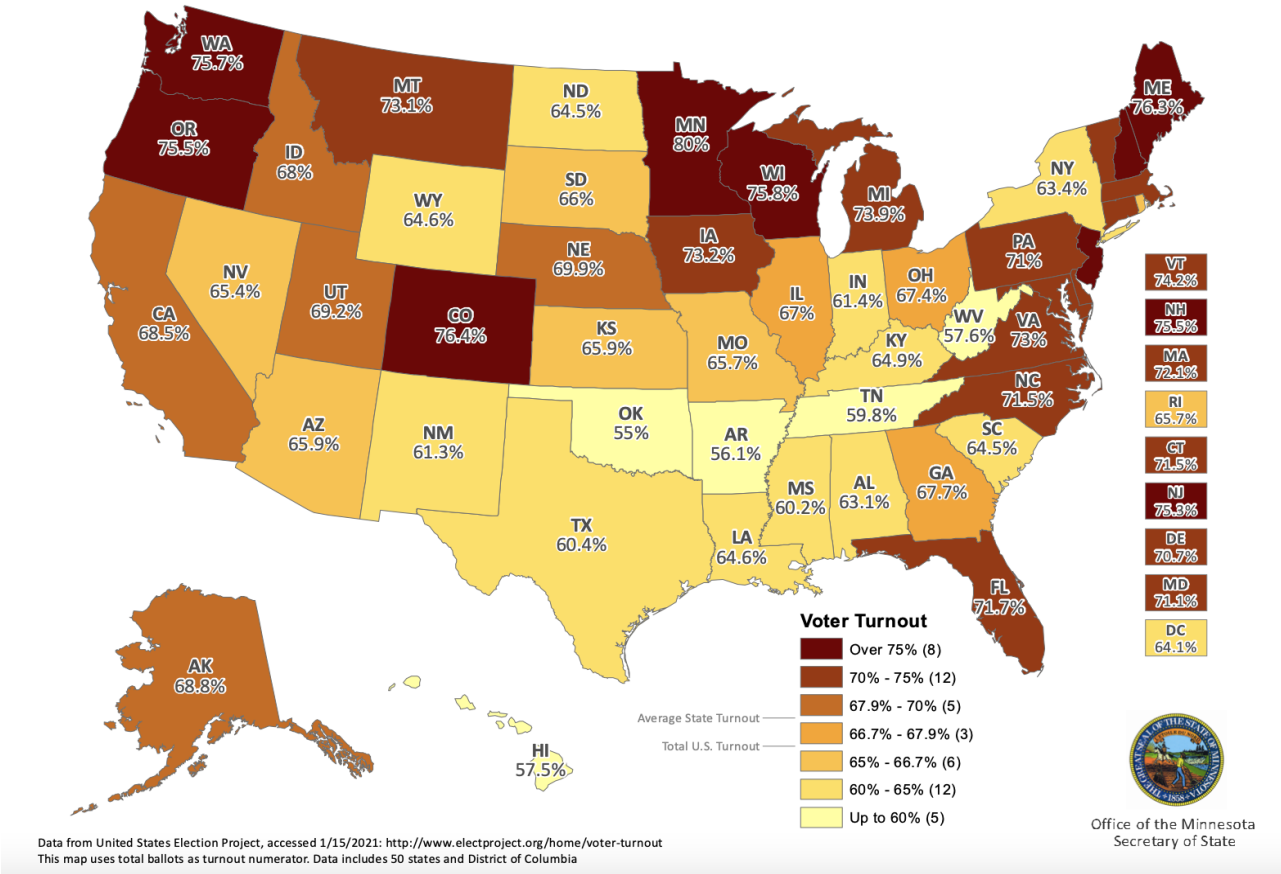
# Research Question

- Last week: Good question is interesting to many people and has broad implications.
- Question identifies problem clearly.
- Not answered by existing research or knowledge.

# Our Research Question - 1

## Turnout

Why do some people choose not to vote?



# Our Research Question - 2

**Turnout**

Why do some states have higher turnout than other states? Which structural factors can explain this variation in voter turnout?

- Voter turnout is relevant in democracies.
- Lots of variation.
- Existing theories were not very strong.

# Theories and Models

## Scientific Modeling

... the generation of a physical, conceptual, or mathematical representation of a real phenomenon that is difficult to observe directly. Scientific models are used to explain and predict the behaviour of real objects or systems and are used in a variety of scientific disciplines, ranging from physics and chemistry to ecology and the Earth sciences. Although modeling is a central component of modern science, scientific models at best are approximations of the objects and systems that they represent – they are not exact replicas. Thus, scientists constantly are working to improve and refine models.

(Source: Encyclopedia Britannica)

# Theories and Models

- Models are not small-scale replications of the real world.
- Models are abstractions capturing relevant features.
  - Geocentric model was *wrong* but you could still build sundial off of it.
- Models, in this sense, are like maps
- We judge models by how useful they are and *not* whether they are a good replica of reality.

# Theories and Models

- Models are not small-scale replications of the real world.
- Models are abstractions capturing relevant features.
  - Geocentric model was *wrong* but you could still build sundial off of it.
- Models, in this sense, are like maps
- We judge models by how useful they are and *not* whether they are a good replica of reality.

"Essentially, all models are wrong, but some are useful."

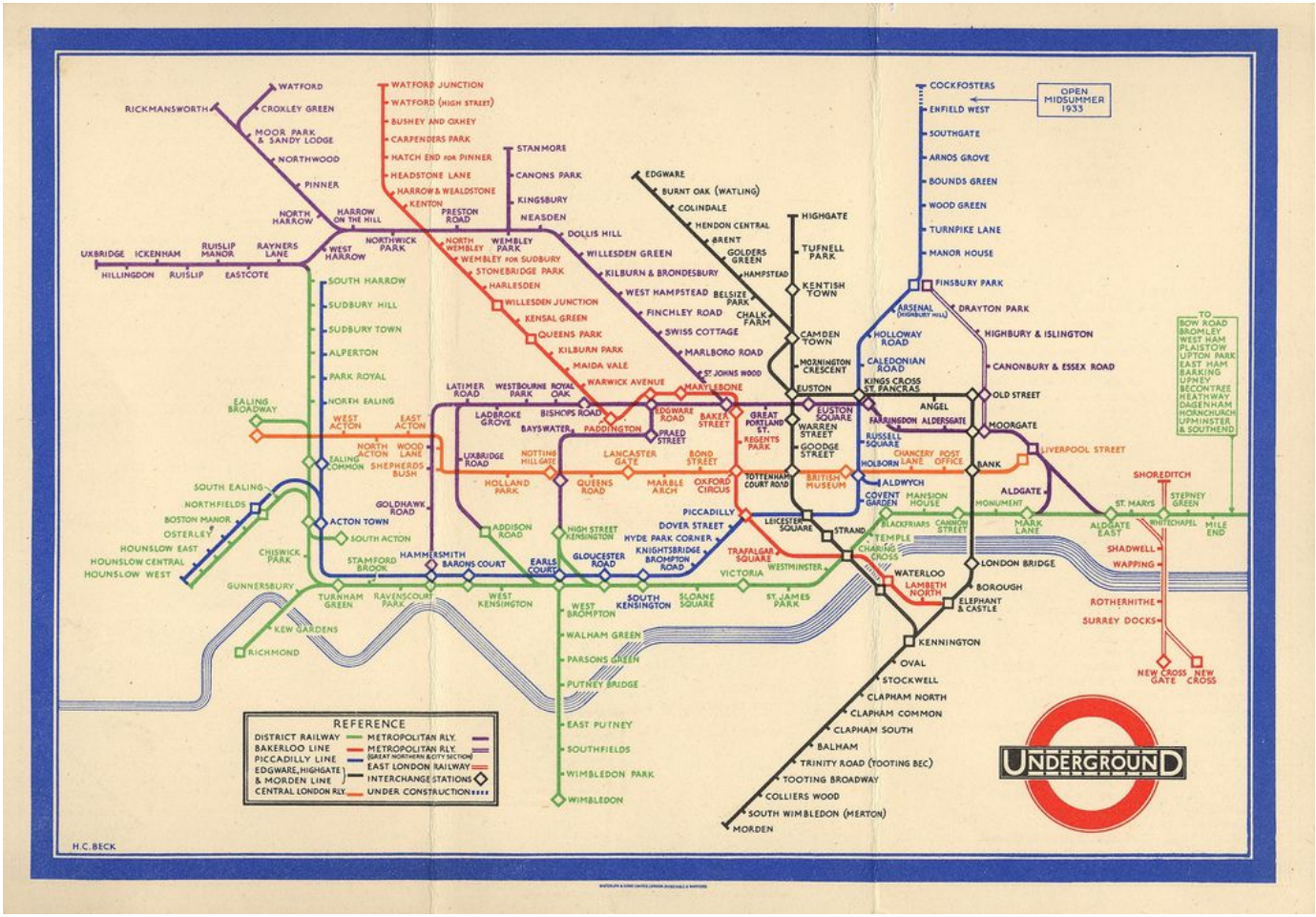
P.E. Box (1987)

# Theories and Models

## Realistic London underground map by 1920

# Theories and Models

Less realistic but more useful map by Harry Beck in 1933



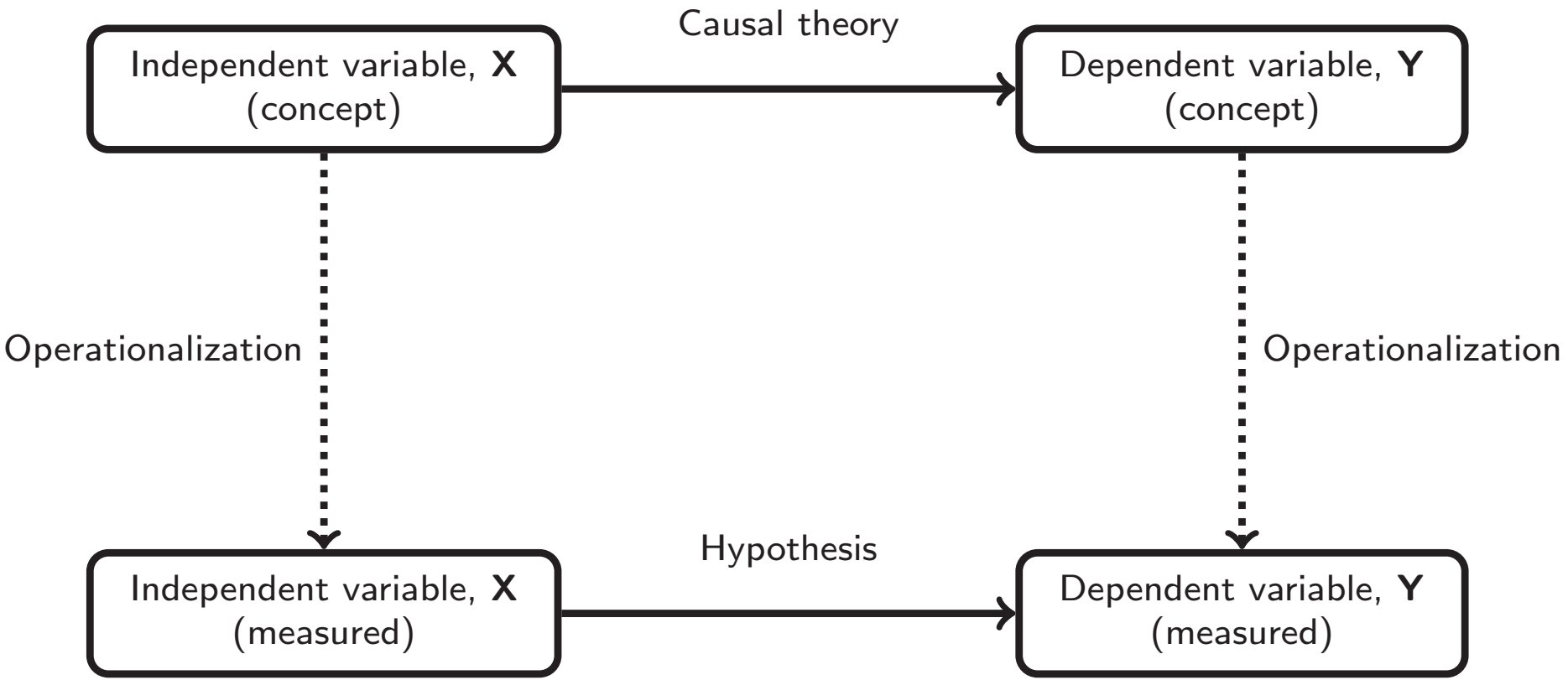
# Example: Rational Voter

Individual utility function:

$$R = p \cdot B - C$$

- $p$: Probability that a voter's vote is decisive, that the vote will yield the preferred outcome.
- $B$: How much more a voter prefers his/her party over the opposite, i.e. utility gain from getting the preferred outcome.
- $C$: Costs of voting.
- $R$: If the benefits outweigh the costs, $R$ **is positive** and an individual is more likely to vote.

# Operationalization and hypotheses



# Causal theory

From our model we predict a number of *probabilistic* causal effects:

- Dependent variable: Turnout (R)
- Independent variables: $p$, $B$, $C$
- Causal effects:
  - $p$: As the probability of casting the decisive vote increases, we should see higher turnout.
  - $B$: For individuals that feel stronger about the consequences of the vote outcome, we should see higher turnout.
  - $C$: In areas where voting is more costly the turnout should be lower.

# Operationalization and hypotheses

- Individuals with strong **party attachments (B)** are more **likely to vote**.
- Where voting is easier (**registration rules (C)**), **turnout** will be higher.
- **Educated people (B)** care more about the election outcomes and are more **likely to vote**.
- **Younger people (B)** have less strong feelings about the outcome of elections.
- People **thinking that voting can make a difference (p)** are more **likely to vote**.

# Data Collection

What data do we need?

- Turnout data from all states.
- Two different time periods.
- We look at how some variables change over time and whether turnout then also changes
  - Registration made easier.
  - Level of education.
  - Average age of voters.
  - ...

# Data: Testing Hypotheses

**Table 4-1** Decomposition of the Decline in Voter Turnout in Presidential Election Years Between the 1960s and 1980s

| Change | Effect on Percentage Change in Turnout Between 1960s and 1980s | Percentage of Decline in Turnout Explained |
|---|---|---|
| An easing of voter registration laws | +1.8 | |
| Increased formal education | +2.8 | |
| A younger electorate | −2.7 | 17 |
| Weakened social involvement | −1.4 | 9 |
| Declining feelings of efficacy | −1.4 | 9 |
| Weakened attachment to and evaluations of the political parties and their candidates | −1.7 | 11 |
| A decline in mobilization | −8.7 | 54 |
| | | 100 |
| Net change in voter turnout: | −11.3 | |

*Source:* Appendix.

Source: Niemi and Weisberg. 2001. *Controversies in Voting Behavior*. New York: CQ Press, p. 71.

# Evaluating the Hypotheses

- We found empirical support for our hypotheses.
- For the moment we might think that the model is a good approximation.
- We can use it to formulate policy recommendations:
  - Educate people about how elections make a difference (affecting $B$).
  - Making participation easy (affecting $C$) or incentivizing it (voter party).

# The Four Hurdles to Assess Causal Statements

1. Is there a credible causal mechanism that connects $X$ to $Y$?
2. Can we rule out the possibility that $Y$ could cause $X$?
3. Is there covariation between $X$ and $Y$?
4. Have we controlled for all confounding variables $Z$ that might make the association between $X$ and $Y$ spurious?

# Examples


1. Is there a credible causal mechanism?
2. Can we rule out: $Y$ cause $X$?
3. Is there an empirical relationship?
4. Do we control for all other variables $Z$?

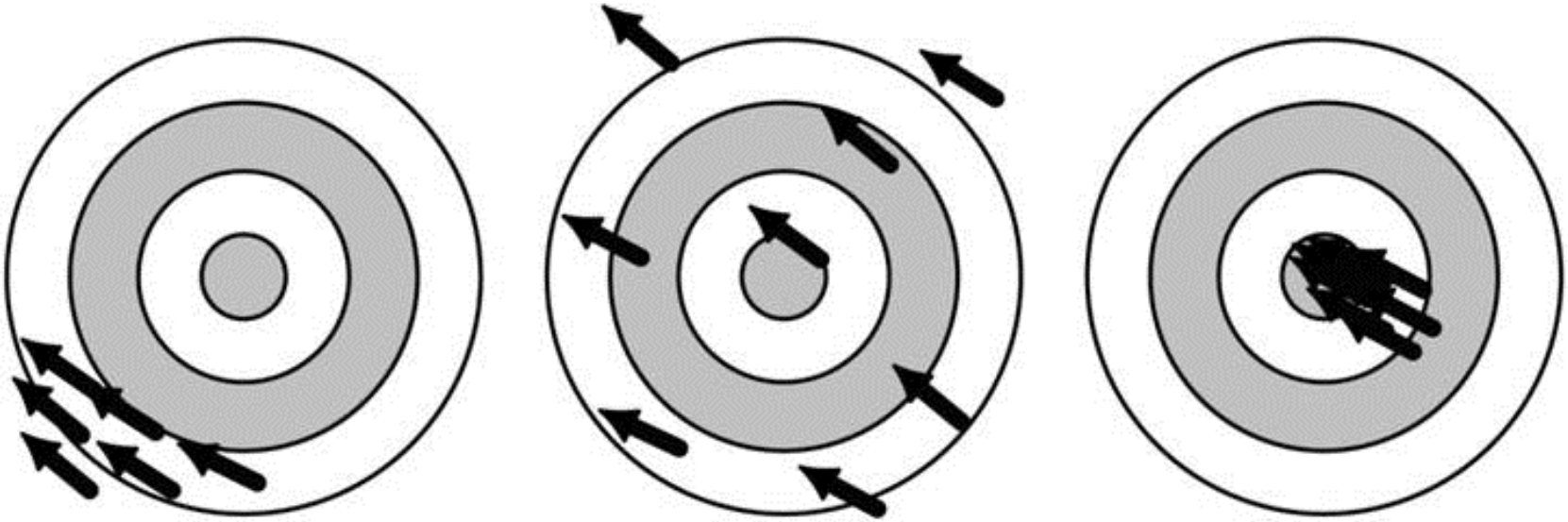
# Validity and Reliability

Are your measures good?

- Validity – am I measuring what I want to measure?
  - Measuring education by asking people if they know what this is: $f(x) = 2 \cdot x^3 + x^2 - 12 \cdot x$ for $x = 2$.
- Reliability – is my measure consistent?
  - "How do you feel today?" as a survey question to measure life satisfaction.

# Validity and Reliability

**FIGURE 5.3** **Comparing the Reliability and Validity of Three Measures**




Source: Clark, Golder, and Golder (2013)

# Good Practice

- Identifying causality is an essential skill
- Doubt and skepticism is key
- Challenge causal claims with the 4 hurdles
- Think formally
- Clearly define: dependent (Y), independent (X), and control/confounder (Z) variables
- Be open about potential pitfalls

# Take-Home Points

- Question -> theory/model -> operationalization/hypotheses -> data collection/test/evaluate
- Models are not picturing reality but highlighting key aspects
- Theory/models predict causal associations between dependent and independent variables
- Dependent and independent variables are operationalized, associations constitute hypotheses
- Hypotheses are tested with data and theory is evaluated
- Causal claims can be evaluated by four simple questions
- Beware the measurement