Session IV
Practical Issues

Thomas J. Leeper

Government Department
London School of Economics and Political Science

How do we find participants?

- Volunteers
  - Volunteer Science
  - In-house subject pool
- Paid crowdworkers
  - Prolific Academic
  - Mechanical Turk
  - Crowdflower
- "Representative" samples
  - Big players: YouGov, TNS, Gallup, Nielsen, GfK
  - Others: Kantar, SSI, Lucid

SUTO Framework

- Cronbach (1986) talks about generalizability in terms of UTO
- Shadish, Cook, and Campbell (2001) speak similarly of:
    - **S**ettings
    - **U**nits
    - **T**reatments
    - **O**utcomes
- External validity depends on all of these

Population

- Setting
- Units
- Treatments
- Outcomes

Your Study

- Setting
- Units
- Treatments
- Outcomes

Population

- Setting
- Units
- Treatments
- Outcomes

Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?

Population

- Setting
- Units
- Treatments
- Outcomes

Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?
how do these differ?

Population

- Setting
- Units
- Treatments
- Outcomes

Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?
how do these differ?
do these differences matter?

Common Differences

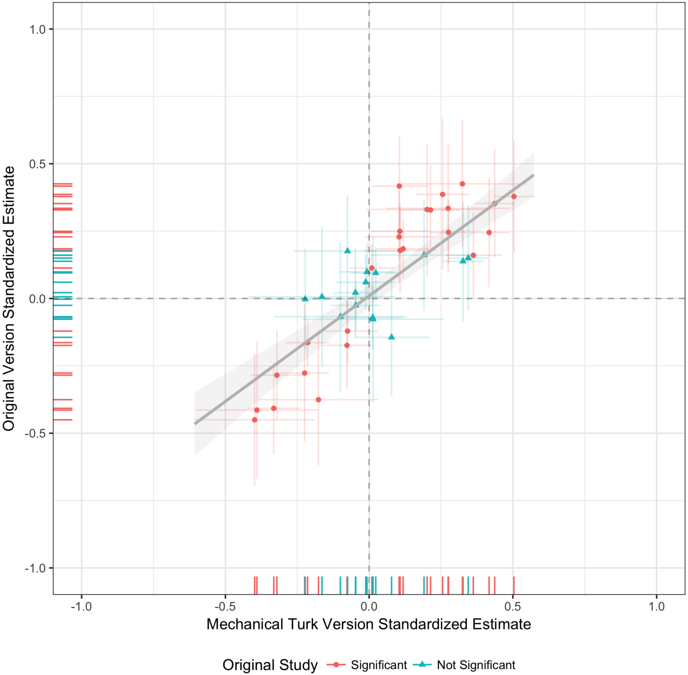
- Most common thing to focus on is demographic representativeness
  - Sears (1986): “students aren’t real people”
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants

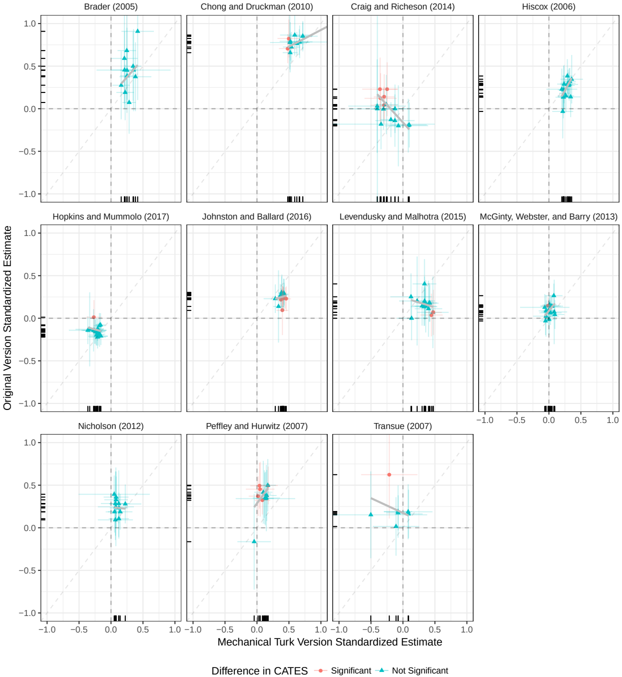
Common Differences

- Most common thing to focus on is demographic representativeness
  - Sears (1986): "students aren't real people"
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants
- But do those characteristics actually matter?

Common Differences

- Most common thing to focus on is demographic representativeness
  - Sears (1986): "students aren't real people"
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants
- But do those characteristics actually matter?
- Shadish, Cook, and Campbell tell us to think about:
  - Surface similarities
  - Ruling out irrelevancies
  - Making discriminations
  - Interpolation/extrapolation


1.0
0.5
0.0
-0.5
-1.0
Original Version Standardized Estimate
-1.0
-0.5
0.0
0.5
1.0
Mechanical Turk Version Standardized Estimate
Original Study
Significant
Not Significant


Brader (2005)
Chong and Druckman (2010)
Craig and Richeson (2014)
Hiscox (2006)
Hopkins and Mummolo (2017)
Johnston and Ballard (2016)
Levendusky and Malhotra (2015)
McGinty, Webster, and Barry (2013)
Nicholson (2012)
Peffley and Hurwitz (2007)
Transue (2007)
Original Version Standardized Estimate
Mechanical Turk Version Standardized Estimate
−1.0
−0.5
0.0
0.5
1.0
Difference in CATES
Significant
Not Significant

Questions?

One final issue with unit-related sources of heterogeneity is how we handle or analyze survey-experimental data where we think participants misbehaved.

One final issue with unit-related sources of heterogeneity is how we handle or analyze survey-experimental data where we think participants misbehaved.

This falls into a couple of broad categories:

- Noncompliance
- Inattention
- Survey Satisficing

How should we deal with respondents that appear to not be paying attention, not "taking" the treatment, or not responding to outcome measures?

1. Keep them
2. Throw them away

Best Practice: Pre-Analysis Protocol

- Excluding respondents based on survey behavior is one of the easiest ways to "p-hack" an experimental dataset
  - Inattention, satisficing, etc. will tend to reduce the size of the SATE
- So regardless of how you handle these respondents, these should be decisions that are made *pre-analysis*

When are you excluding participants?

Pre-Treatment	Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment
- “Failing” a manipulation check

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment
- "Failing" a manipulation check
- Drop-off

Pre-Treatment Exclusion

- This is totally fine from a causal inference perspective

Pre-Treatment Exclusion

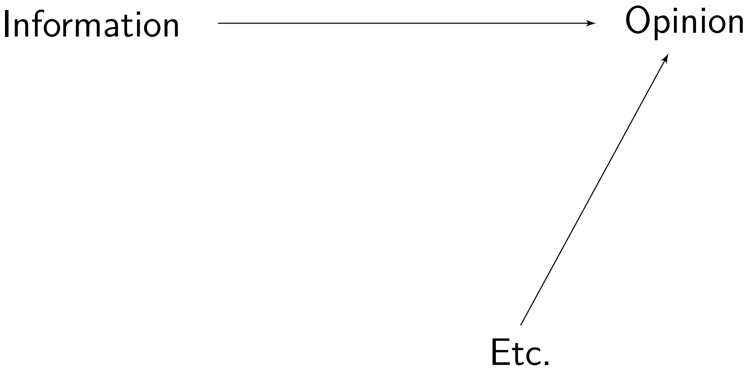
- This is totally fine from a causal inference perspective
- Advantages:
  - Focused on engaged respondents
  - Likely increase impact of treatment

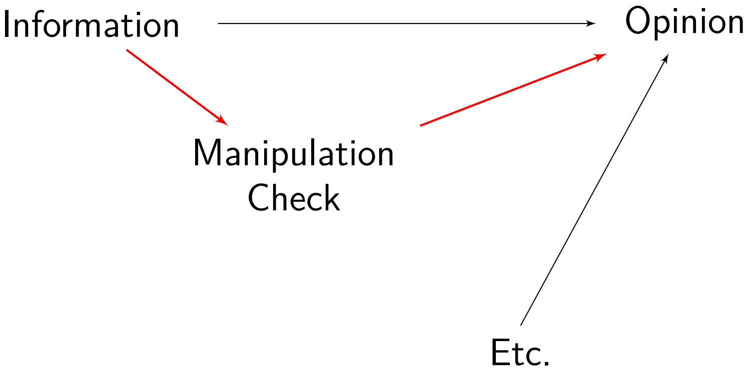
Pre-Treatment Exclusion

- This is totally fine from a causal inference perspective
- Advantages:
  - Focused on engaged respondents
  - Likely increase impact of treatment
- Disadvantages:
  - Changing definition of sample (and thus population)

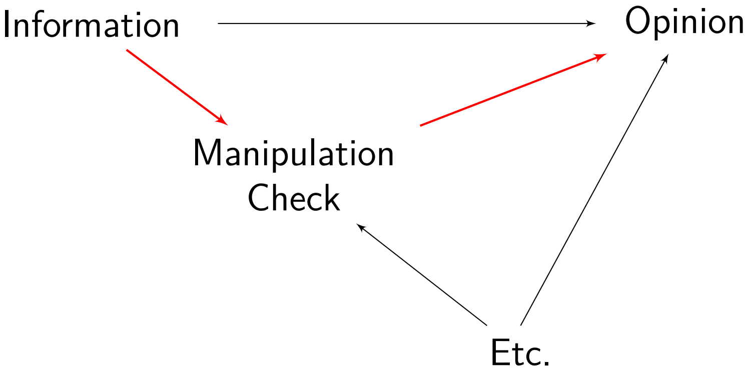
Post-Treatment Exclusion

This is much more problematic because it involves controlling for a *post-treatment* variable


Information
Opinion
Etc.



Risk that estimate of β_1 is diminished because effect is being carried through the manipulation check.



Introduction of "collider bias" wherein values of the manipulation check are affected by other factors.

Post-Treatment Exclusion

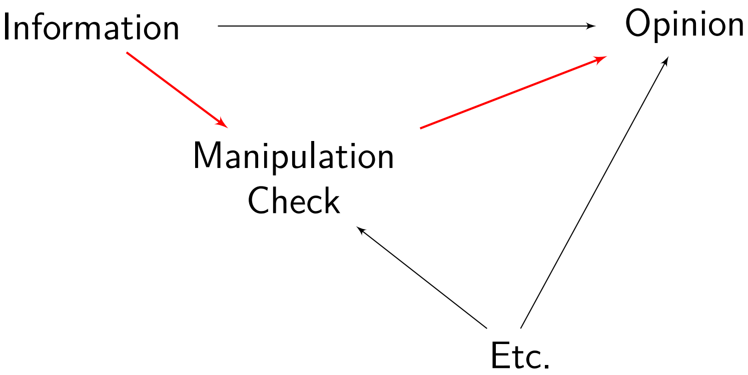
- Any post-treatment exclusion is problematic and should be avoided

Post-Treatment Exclusion

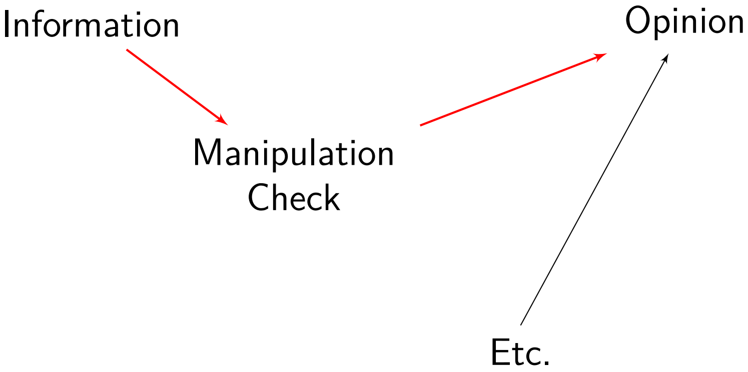
- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation
- Non-response or attrition is the same as researcher-imposed exclusion
  - Not problematic if MCAR
  - Nothing really to be done if caused by treatment



Introduction of "collider bias" wherein values of the manipulation check are affected by other factors.


Information
Manipulation Check
Opinion
Etc.

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
    - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation
- Non-response or attrition is the same as researcher-imposed exclusion
    - Not problematic if MCAR
    - Nothing really to be done if caused by treatment

Questions?

Apparent Satisficing

- Some common measures:
  - "Straightlining"
  - Non-differentiation
  - Acquiescence
  - Nonresponse
  - DK responding
  - Speeding
- Difficult to detect and distinguish from "real" responses

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in
- Record focus/blur browser events

Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?

Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?
- While taking this survey, did you engage in any of the following behaviors? Please check all that apply.
    - Use your mobile phone
    - Browse the internet
    - …

Instructional Manipulation Check

We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click “next” to proceed with the survey.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Instructional Manipulation Check

Do you agree or disagree with the decision to send British forces to fight ISIL in Syria? We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click "next" to proceed with the survey.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Return

Treatment Noncompliance

- Definition:
  "when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated" [1]

[1] Gerber & Green. 2012. *Field Experiments*, p.132.

Treatment Noncompliance

- Definition:
  "when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated" [1]
- Several strategies
  - "As treated" analysis
  - "Intention to treat" analysis
  - Estimate a LATE

[1] Gerber & Green. 2012. *Field Experiments*, p.132.

Analyzing Noncompliance

- If noncompliance only occurs in one group, it is *asymmetric* or *one-sided*
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned: $ITT = \overline{Y}_1 - \overline{Y}_0$

Analyzing Noncompliance

- If noncompliance only occurs in one group, it is *asymmetric* or *one-sided*
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned: $ITT = \overline{Y}_1 - \overline{Y}_0$
- We can use "instrumental variables" to estimate the "local average treatment effect" (LATE) for those that complied with treatment: $LATE = \frac{ITT}{\%Compliant}$

Local Average Treatment Effect

- IV estimate is *local* to the variation in X that is due to variation in D
- This matters if effects are *heterogeneous*
- LATE is effect for those who *comply*
- Four subpopulations:
  - Compliers: $X = 1$ only if $D = 1$
  - Always-takers: $X = 1$ regardless of D
  - Never-takers: $X = 0$ regardless of D
  - Defiers: $X = 1$ only if $D = 0$
- Exclusion restriction! Monotonicity!

Questions?

Discussion

Consider the following:

- When **are we required to** include covariates in the analysis of an experiment?
- When **are we allowed to** include covariates in the analysis of an experiment?
- When **are we not allowed to** include covariates in the analysis of an experiment?

Discuss with a partner for 2 minutes.

- We never have to use covariates!

- We never have to use covariates!
- We may want to for:
  - Subgroup comparisons
  - Repeated/panel designs
  - In case of noncompliance or attrition

- We never have to use covariates!
- We may want to for:
  - Subgroup comparisons
  - Repeated/panel designs
  - In case of noncompliance or attrition
- Any use of covariates should be planned!

Block Randomization I

Stratification:Sampling::Blocking:Experiments

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment
- CATE is estimate for each stratum; aggregated to SATE

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment
- CATE is estimate for each stratum; aggregated to SATE
- Why?
  - Eliminate chance imbalances
  - Optimized for estimating CATEs
  - More precise SATE estimate

Exp.	Control				Treatment			
1	M	M	M	M	F	F	F	F
2	M	M	M	F	M	F	F	F
3	M	M	F	F	M	M	F	F
4	M	F	F	F	M	M	M	F
5	F	F	F	F	M	M	M	M

```
# population of men and women
pop <- rep(c("Male", "Female"), each = 4)

# randomly assign into treatment and control
split(sample(pop, 8, FALSE), c(rep(0,4), rep(1,4)))
```

Obs.	X_{1i}	X_{2i}	D_i
1	Male	Old	0
2	Male	Old	1
3	Male	Young	1
4	Male	Young	0
5	Female	Old	1
6	Female	Old	0
7	Female	Young	0
8	Female	Young	1

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
  - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
  - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE
  - Most valuable in small samples

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
    - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE
    - Most valuable in small samples
    - Not valuable if all blocks have similar potential outcomes

Statistical Properties I

Complete randomization:

$$SATE = \frac{1}{n_1}\sum Y_{1i} - \frac{1}{n_0}\sum Y_{0i}$$

Block randomization:

$$SATE_{blocked} = \sum_{1}^{J}\left(\frac{n_j}{n}\right)(\widehat{CATE}_j)$$

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	
2	Male	Old	1	10	
3	Male	Young	1	4	
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	4
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	4
6	Female	Old	0	2	
7	Female	Young	0	6	3
8	Female	Young	1	9	

SATE Estimation

$$
\begin{aligned}
SATE &= \left(\frac{2}{8} * 5\right) + \left(\frac{2}{8} * 3\right) + \left(\frac{2}{8} * 4\right) + \left(\frac{2}{8} * 3\right) \\
&= 3.75
\end{aligned}
$$

SATE Estimation

$$
\begin{aligned}
SATE &= \left(\frac{2}{8} * 5\right) + \left(\frac{2}{8} * 3\right) + \left(\frac{2}{8} * 4\right) + \left(\frac{2}{8} * 3\right) \\
&= 3.75
\end{aligned}
$$

The blocked and unblocked estimates are the same here because $Pr(Treatment)$ is constant across blocks and blocks are all the same size.

SATE Estimation

- We can use weighted regression to estimate this in an OLS framework
- Weights are the inverse prob. of being treated w/in block
  - Pr(Treated) by block: $p_{ij} = Pr(D_i = 1 | J = j)$
  - Weight (Treated): $w_{ij} = \frac{1}{p_{ij}}$
  - Weight (Control): $w_{ij} = \frac{1}{1 - p_{ij}}$

Statistical Properties II

Complete randomization:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{n_0} + \frac{\widehat{Var}(Y_1)}{n_1}}$$

Block randomization:

$$\widehat{SE}_{SATE_{blocked}} = \sqrt{\sum_{1}^{J} \left(\frac{n_j}{n}\right)^2 \widehat{Var}(SATE_j)}$$

Statistical Properties II

Complete randomization:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{n_0} + \frac{\widehat{Var}(Y_1)}{n_1}}$$

Block randomization:

$$\widehat{SE}_{SATE_{blocked}} = \sqrt{\sum_{1}^{J} \left(\frac{n_j}{n}\right)^2 \widehat{Var}(SATE_j)}$$

When is the blocked design more efficient?

Practicalities

- Blocked randomization only works in exactly the same situations where stratified sampling works
  - Need to observe covariates pre-treatment in order to block on them
  - Work best in a panel context
- In a single cross-sectional design that might be challenging
  - Some software can block "on the fly"

Questions?

Detecting Effect Heterogeneity

Always block if you expect heterogeneity!

- QQ-plots: Suggestive evidence
- Regression using treatment-by-covariate interactions

Detecting Effect Heterogeneity

Always block if you expect heterogeneity!

- QQ-plots: Suggestive evidence
- Regression using treatment-by-covariate interactions
- (Replication and meta-analysis)

Suggestive Evidence

We can never know $Var(TE_i)$!

Suggestive Evidence

We can never know $Var(TE_i)$! But...

- Quantile-quantile plots

Suggestive Evidence

We can never know $Var(TE_i)$! But...

- Quantile-quantile plots
  - Compare the distribution of Y_0's to distribution of Y_1's
  - If homogeneity, a vertical shift in Y_1's
  - If heterogeneity, a slope $\neq 1$

Suggestive Evidence

We can never know $Var(TE_i)$! But...

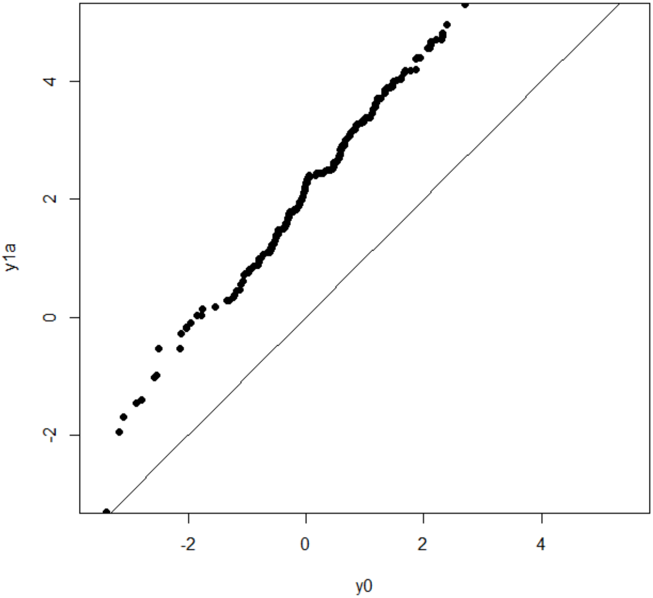
- Quantile-quantile plots
  - Compare the distribution of Y_0's to distribution of Y_1's
  - If homogeneity, a vertical shift in Y_1's
  - If heterogeneity, a slope $\neq 1$
- Equality of variance tests
  - If homogeneity, variance should be equal
  - If heterogeneity, variances should differ

QQ Plots

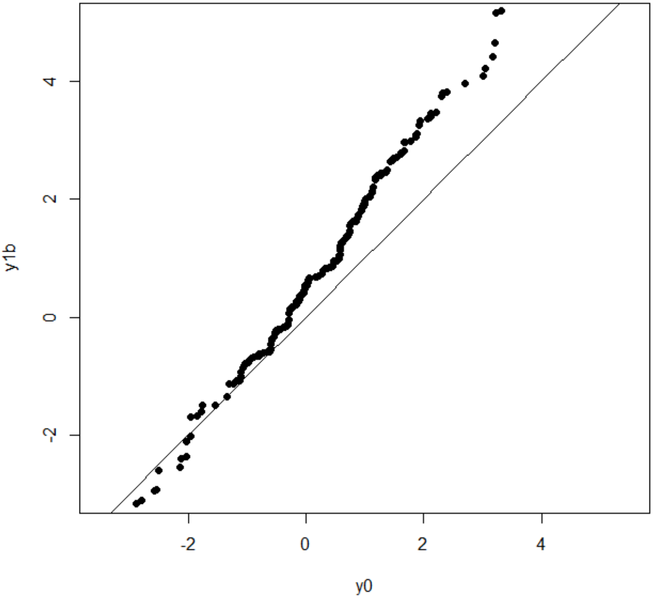
```
# y_0 data
set.seed(1)
n <- 200
y0 <- rnorm(n) + rnorm(n, 0.2)

# y_1 data (homogeneous effects)
y1a <- y0 + 2 + rnorm(n, 0.2)
# y_1 data (heterogeneous effects)
y1b <- y0 + rep(0:1, each = n/2) + rnorm(n, 0.2)

qqplot(y0, y1a, pch=19, xlim=c(-3,5), ylim=c(-3,5), asp=1)
curve((x), add = TRUE)
qqplot(y0, y1b, pch=19, xlim=c(-3,5), ylim=c(-3,5), asp=1)
curve((x), add = TRUE)
```



y1a
-2
0
2
4
y0
-2
0
2
4


y1b
y0
-2
0
2
4

Equality of Variance tests

```
> var.test(y0, y1a)

        F test to compare two variances

data:  y0 and y1a
F = 0.60121, num df = 199, denom df = 199,
  p-value = 0.0003635
alternative hypothesis:
  true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4549900 0.7944289
sample estimates:
ratio of variances
         0.6012131
```

Equality of Variance tests

```
> var.test(y0, y1b)

        F test to compare two variances

data:  y0 and y1b
F = 0.53483, num df = 199, denom df = 199,
  p-value = 1.224e-05
alternative hypothesis:
  true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4047531 0.7067133
sample estimates:
ratio of variances
         0.5348312
```

Questions?

Regression Estimation

Aside: Regression Adjustment in Experiments, Generally

- Recall the general advice that we do not need covariates in the regression to "control" for omitted variables (because there are none)
- Including covariates can reduce variance of our SATE by explaining more of the variation in Y

Scenario

Imagine two regression models. Which is correct?

1. Mean-difference estimate of SATE is "not significant"
2. Regression estimate of SATE, controlling for sex, age, and education, is "significant"

Scenario

Imagine two regression models. Which is correct?

1. Mean-difference estimate of SATE is "not significant"
2. Regression estimate of SATE, controlling for sex, age, and education, is "significant"

This is a small-sample dynamic, so make these decisions pre-analysis!

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$
- CATEs:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 X * M + e$$

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$
- CATEs:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 X * M + e$$

  - Homogeneity: $\beta_3 = 0$
  - Heterogeneity: $\beta_3 \neq 0$

Questions?

Quiz time!

Compliance

1. What is compliance?

Compliance

1. What is compliance?
2. How can we analyze experimental data when there is noncompliance?

Balance testing

1. What does randomization ensure about the composition of treatment groups?

Balance testing

1. What does randomization ensure about the composition of treatment groups?
2. What can we do if we find a covariate imbalance between groups?

Balance testing

1. What does randomization ensure about the composition of treatment groups?
2. What can we do if we find a covariate imbalance between groups?
3. How can we avoid this problem entirely?

Nonresponse and Attrition

1. Do we care about outcome nonresponse in experiments?

Nonresponse and Attrition

1. Do we care about outcome nonresponse in experiments?
2. How can we analyze experimental data when there is outcome nonresponse or post-treatment attrition?

Manipulation checks

1. What is a manipulation check? What can we do with it?

Manipulation checks

1. What is a manipulation check? What can we do with it?
2. What do we do if some respondents "fail" a manipulation check?

Null effects

1. What should we do if we find our estimated $\widehat{SATE} = 0$?

Null effects

1. What should we do if we find our estimated $\widehat{SATE} = 0$?
2. What does it mean for an experiment to be *underpowered*?

Null effects

1. What should we do if we find our estimated $\widehat{SATE} = 0$?
2. What does it mean for an experiment to be *underpowered*?
3. What can we do to reduce the probability of obtaining an (unwanted) “null effect”?

Effect heterogeneity

1. What should we do if, post-hoc, we find evidence of effect heterogeneity?

Effect heterogeneity

1. What should we do if, post-hoc, we find evidence of effect heterogeneity?
2. What can we do pre-implementation to address possible heterogeneity?

Representativeness

1. Under what conditions is a design-based, probability sample necessary for experimental inference?

Representativeness

1. Under what conditions is a design-based, probability sample necessary for experimental inference?
2. What kind of causal inferences can we draw from an experiment on a descriptively unrepresentative sample?

Peer Review

1. What should we do if a peer reviewer asks us to “control” for covariates in the analysis?

Peer Review

1. What should we do if a peer reviewer asks us to "control" for covariates in the analysis?
2. What should we do if a peer reviewer asks us to include or exclude particular respondents from the analysis?

Questions?

History: Key Moments

1. Tuskegee (1932-1972) and Guatemala (1946-1948) Studies
2. Nuremberg Code (1947)
3. Helsinki Declaration (1964)
4. U.S. 45 CFR 46 (1974) and “Common Rule” (1991)
5. The Belmont Report (1979)
6. EU Data Protection Directive (1995; 2012)
    - UK Data Protection Act (1998)

Helsinki Declaration

- Adopted by the World Medical Association in 1964[2]
- Narrowly focused on medical research
- Expanded the Nuremberg Code
  - Relaxed consent requirements
  - Risks should not exceed benefits
  - Institutionalization of ethics oversight

[2] http://www.bmj.com/content/2/5402/177

Helsinki Declaration

- Adopted by the World Medical Association in 1964[2]
- Narrowly focused on medical research
- Expanded the Nuremberg Code
    - Relaxed consent requirements
    - Risks should not exceed benefits
    - Institutionalization of ethics oversight
- Do these rules apply to non-medical research?

[2] http://www.bmj.com/content/2/5402/177

The Belmont Report

- Commissioned by the U.S. Government in 1979[3]
- Three overarching principles:
  1. Respect for persons
  2. Beneficence
  3. Justice
- Three policy implications:
  - Informed consent
  - Assessment of risks/benefits
  - Care for vulnerable populations

[3] http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Benefits and Harm

- What is a “benefit”?
- What is a “harm”?
- How do we balance the two?

Ethical Considerations

- Most ethical issues are not unique to *experimental* social science
- Some especially important issues:
    1. Randomization
    2. Informed consent
    3. Privacy
    4. Deception
    5. Publication bias

I. Randomization

- Is it ethical to randomize?

II. Informed Consent

- Persons must consent to being a research subject

II. Informed Consent

- Persons must consent to being a research subject
- What this means in practice is complicated
  - What is consent?
  - What is "informed" consent?
  - What exactly do they have to consent to?

II. Informed Consent

- Persons must consent to being a research subject
- What this means in practice is complicated
    - What is consent?
    - What is "informed" consent?
    - What exactly do they have to consent to?
- Cross-national variations
    - Consent forms required in U.S.
    - Not required in UK

III. Privacy

- Under EU Data Protection Directive (1995), data can be processed when:
  - Consent is given
  - Data are used for a "legitimate" purpose
  - Anonymous or confidential
- These rules have become more expansive under GDPR (in force as of 2018)
- Data cannot leave the EU except under conditions

III. Privacy

- Experimental might be additionally sensitive

III. Privacy

- Experimental might be additionally sensitive
- Answers reflect "manipulated" attitudes, behaviors, perceptions, etc. that respondents may not have given in another setting

IV. Deception

- Major distinction between psychology tradition and economics tradition[4]
  - Purpose of the study
  - Purpose of specific items or tasks
  - Order or length of questionnaire

[4] Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition[4]
  - Purpose of the study
  - Purpose of specific items or tasks
  - Order or length of questionnaire
- Psychologists focus on *debriefing*

[4]Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition[4]
  - Purpose of the study
  - Purpose of specific items or tasks
  - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*

[4] Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition[4]
  - Purpose of the study
  - Purpose of specific items or tasks
  - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*
  - Omission: In a multi-round trust game, an additional round is added

[4] Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition[4]
  - Purpose of the study
  - Purpose of specific items or tasks
  - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*
  - Omission: In a multi-round trust game, an additional round is added
  - Commission: Telling respondents it is a dictator game, but it is actually a trust game

[4] Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

V. Publication Bias

- Publication bias not typically discussed as an ethical question

V. Publication Bias

- Publication bias not typically discussed as an ethical question
- If studies are meant to policy or practical implications, then we care about PATE or a set of CATEs, including whether their effects are positive, negative, or zero.

V. Publication Bias

- Publication bias not typically discussed as an ethical question
- If studies are meant to policy or practical implications, then we care about PATE or a set of CATEs, including whether their effects are positive, negative, or zero.
- Publication bias (toward "significant" results) invites wasting resources on treatments that actually don't work

Lots of Other Ethical Questions

Lots of Other Ethical Questions

1. Funding

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization
3. Vulnerable populations (e.g. children, sick)

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization
3. Vulnerable populations (e.g. children, sick)
4. Incentives

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization
3. Vulnerable populations (e.g. children, sick)
4. Incentives
5. Cross-national research

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization
3. Vulnerable populations (e.g. children, sick)
4. Incentives
5. Cross-national research
6. End uses/users of research

Lots of Other Ethical Questions

1. Funding
2. Independence and Politicization
3. Vulnerable populations (e.g. children, sick)
4. Incentives
5. Cross-national research
6. End uses/users of research
7. Others. . .

Questions?

Learning Outcomes

By the end of the week, you should be able to. . .

Learning Outcomes

By the end of the week, you should be able to. . .

1. Explain how to analyze experiments quantitatively.

Learning Outcomes

By the end of the week, you should be able to. . .

1. Explain how to analyze experiments quantitatively.
2. Explain how to design experiments that speak to relevant research questions and theories.

Learning Outcomes

By the end of the week, you should be able to. . .

1. Explain how to analyze experiments quantitatively.
2. Explain how to design experiments that speak to relevant research questions and theories.
3. Evaluate the uses and limitations of several common survey experimental paradigms.

Learning Outcomes

By the end of the week, you should be able to. . .

1. Explain how to analyze experiments quantitatively.
2. Explain how to design experiments that speak to relevant research questions and theories.
3. Evaluate the uses and limitations of several common survey experimental paradigms.
4. Identify practical issues that arise in the implementation of experiments and evaluate how to anticipate and respond to them.

Wrap-up

- Thanks to all of you!
- Stay in touch (t.leeper@lse.ac.uk)
- Good luck with your research!

5 Beyond One-Shot Designs

6 Behavioral Outcomes

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
  1. Measure effect duration in two-wave panel
  2. Solicit pre-treatment outcome measures in a two-wave panel
  3. Measure effects of field treatment in post-test only design
  4. Randomly encourage field treatment in pre-test and measure effects in post-test

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
  1. Measure effect duration in two-wave panel
  2. Solicit pre-treatment outcome measures in a two-wave panel
  3. Measure effects of field treatment in post-test only design
  4. Randomly encourage field treatment in pre-test and measure effects in post-test
- Problems? Compliance & nonresponse

I. Effect Duration

- Use a two- (or more-) wave panel to measure duration of effects
  - T1: Treatment and outcome measurement
  - T2+: Outcome measurement
- Two main concerns
  - Attrition
  - Panel conditioning

II. Within-Subjects Designs

- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

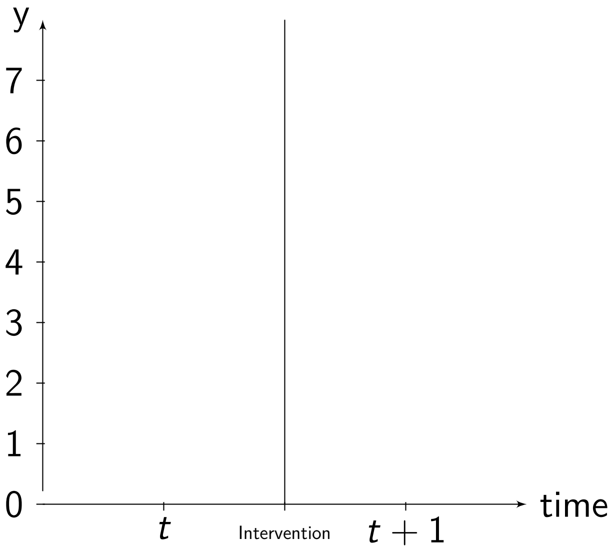
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

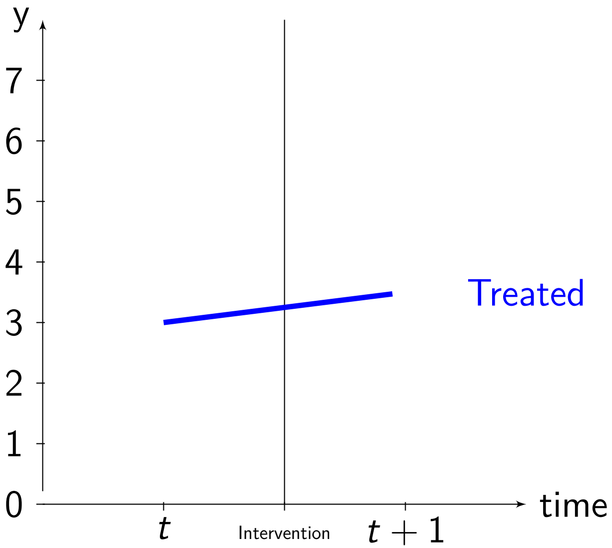
II. Within-Subjects Designs

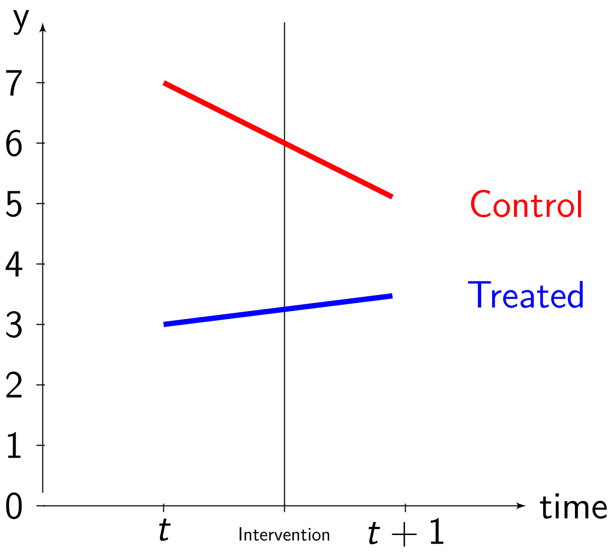
- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

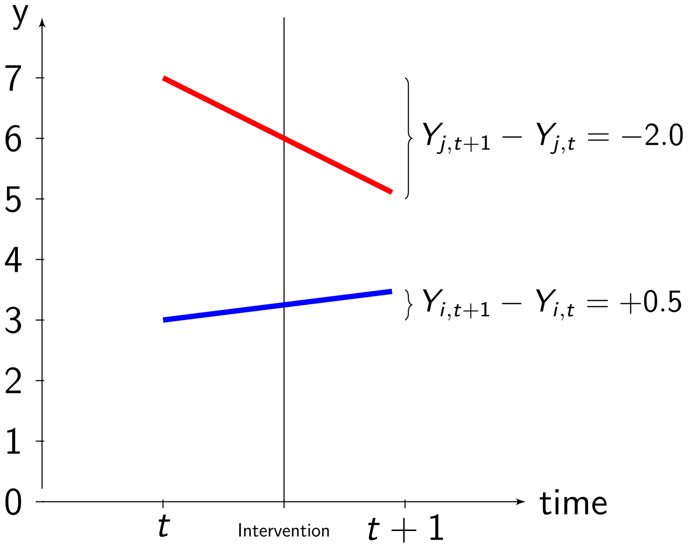
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

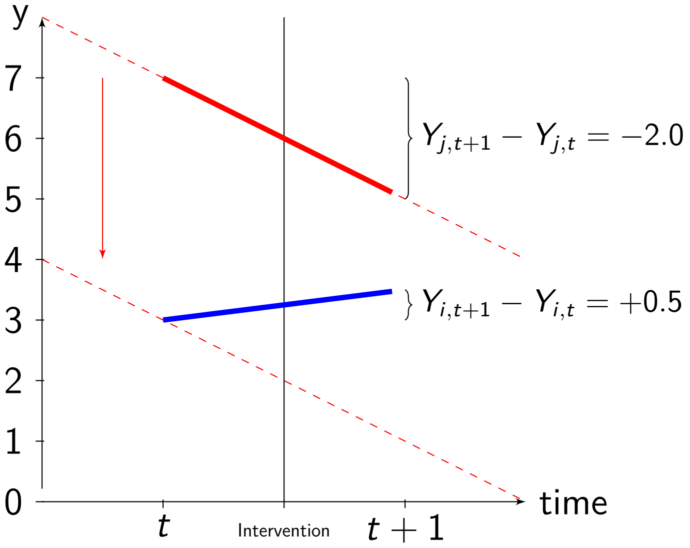
- Advantageous because variance for paired samples decreases as correlation between t_0 and t_1 observations increases

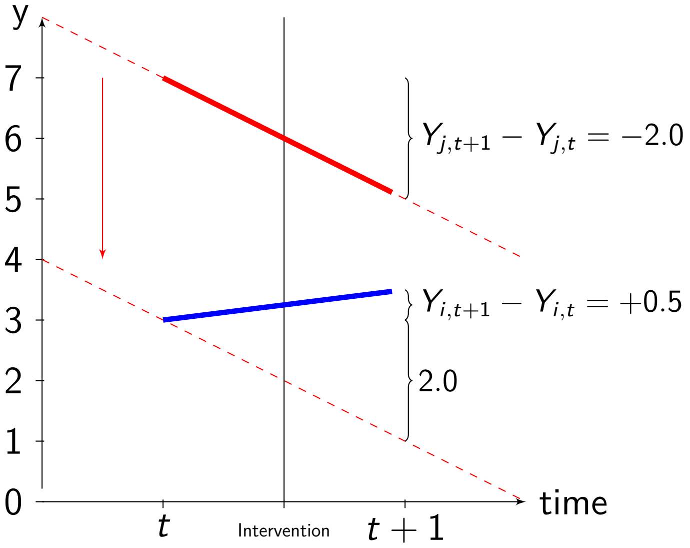

y
7
6
5
4
3
2
1
0
time
t
Intervention
t + 1

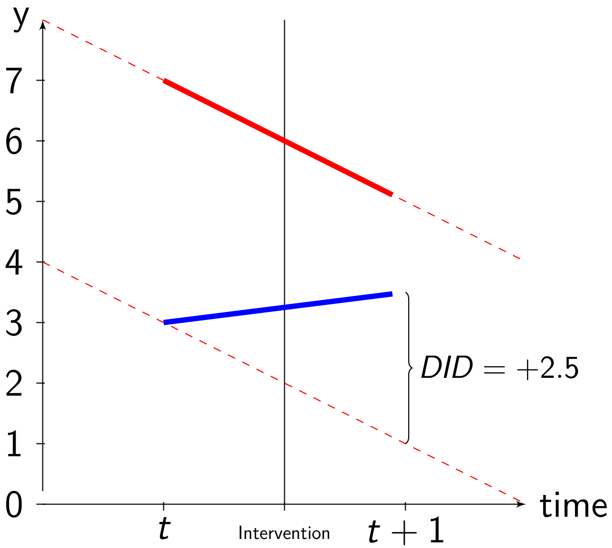

y
7
6
5
4
3
2
1
0
Treated
time
t
Intervention
$t+1$


y
7
6
5
4
3
2
1
0
Control
Treated
time
t
Intervention
$t+1$


y
7
6
5
4
3
2
1
0
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$
time
t
Intervention
$t+1$


y
7
6
5
4
3
2
1
0
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$
time
t
Intervention
$t + 1$


y
7
6
5
4
3
2
1
0
time
t
Intervention
$t+1$
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$
2.0


y
7
6
5
4
3
2
1
0
$DID = +2.5$
time
t
Intervention
$t+1$

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)
2. Maturation (time trends)

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)
5. Instability (measurement error)

[5] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[5]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)
5. Instability (measurement error)
6. Attrition

[5] Shadish, Cook, and Campbell (2002)

III. Randomized Field Treatment

- Examples:

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues
  - Nonresponse
  - Noncompliance

Noncompliance

- Compliance is when individuals receive and accept the treatment to which they are assigned
- Noncompliance:
  "when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated" [6]
- This causes problems for our analysis because factors other than randomization explain why individuals receive their treatment
- Lots of methods for dealing with this, but the consequence is generally reduced power

[6] Gerber & Green. 2012. *Field Experiments*, p.132.

Asymmetric Noncompliance

- Noncompliance *asymmetric* if only in one group
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned $ITT = \overline{Y}_1 - \overline{Y}_0$
- We can use "instrumental variables" to estimate the "local average treatment effect" (LATE) for those that complied with treatment: $LATE = \frac{ITT}{PercentCompliant}$
- We can ignore randomization and analyze data "as-treated", but this makes our study no longer an experiment

Local Average Treatment Effect

- IV estimate is *local* to the variation in X that is due to variation in D
- LATE is effect for those who *comply*
- Four subpopulations:
  - Compliers: $X = 1$ only if $D = 1$
  - Always-takers: $X = 1$ regardless of D
  - Never-takers: $X = 0$ regardless of D
  - Defiers: $X = 1$ only if $D = 0$
- Exclusion restriction! Monotonicity!

Two-Sided Noncompliance

- Two-sided noncompliance is more complex analytically
- Stronger assumptions are required to analyze it and we won't discus them here
- Best to try to develop a better design to avoid this rather than try to deal with the complexities of analyzing a broken design

IV. Treatment Encouragement

- Design:
  - T1: Encourage treatment
  - T2: Measure effects
- Examples:
  1. Albertson and Lawrence[7]

[7] Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300. 10.1177/1532673X08328600.

IV. Treatment Encouragement

- Design:
  - T1: Encourage treatment
  - T2: Measure effects
- Examples:
  1. Albertson and Lawrence[7]
- Issues

[7] Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300. 10.1177/1532673X08328600.

IV. Treatment Encouragement

- Design:
    - T1: Encourage treatment
    - T2: Measure effects
- Examples:
    1. Albertson and Lawrence[7]
- Issues
    - Nonresponse
    - Noncompliance

[7] Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300. 10.1177/1532673X08328600.

Treatment Noncompliance

Treatment Noncompliance

- Several strategies
  - "As treated" analysis
  - "Intention to treat" analysis
  - Estimate a LATE

Questions?

Heterogeneity due to *O*utcomes

- This is expected!
  - E.g., non-equivalent outcomes
- Reasonable to explore multiple outcomes
  - Multiple comparisons
  - Power considerations
  - Construct validity

Heterogeneity due to *O*utcomes

- This is expected!
  - E.g., non-equivalent outcomes
- Reasonable to explore multiple outcomes
  - Multiple comparisons
  - Power considerations
  - Construct validity
- What outcomes you measure depend on your theory

Heterogeneity due to *O*utcomes

- This is expected!
  - E.g., non-equivalent outcomes
- Reasonable to explore multiple outcomes
  - Multiple comparisons
  - Power considerations
  - Construct validity
- What outcomes you measure depend on your theory
- Lots of potential for behavioral measures!

Behavioural measures

Some behaviours that can be directly measured through survey questionnaires.

Behavioural measures

Some behaviours that can be directly measured through survey questionnaires.

Three broad categories:

Behavioural measures

Some behaviours that can be directly measured through survey questionnaires.

Three broad categories:

1. Behavioural measures that provide survey paradata

Behavioural measures

Some behaviours that can be directly measured through survey questionnaires.

Three broad categories:

1. Behavioural measures that provide survey paradata
2. Behavioural measures that operationalize attitudes

Behavioural measures

Some behaviours that can be directly measured through survey questionnaires.

Three broad categories:

1. Behavioural measures that provide survey paradata
2. Behavioural measures that operationalize attitudes
3. Behavioural measures that operationalize behaviours

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies
- Reading times

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies
- Reading times
- Answer switching

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies
- Reading times
- Answer switching
- Eye tracking

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies
- Reading times
- Answer switching
- Eye tracking
- Mouse tracking

Behavioural Measures for Paradata

Why?

- Respondents use of the survey tells us something meaningful about their behaviour

What?

- Nonresponse
- Response latencies
- Reading times
- Answer switching
- Eye tracking
- Mouse tracking
- Smartphone metadata

Behavioural Measures for Attitudes

Why?

- Attitudinal self-reports might be "cheap talk"

Behavioural Measures for Attitudes

Why?

- Attitudinal self-reports might be “cheap talk”

What?

Behavioural Measures for Attitudes

Why?

- Attitudinal self-reports might be "cheap talk"

What?

- Implicit Association Test

Behavioural Measures for Attitudes

Why?

- Attitudinal self-reports might be “cheap talk”

What?

- Implicit Association Test
- Incentivized Survey questions

Behavioural Measures for Behaviour

Why?

- We want to observe or affect behaviour (e.g., in an experiment)

Behavioural Measures for Behaviour

Why?

- We want to observe or affect behaviour (e.g., in an experiment)

What?

- Directly measure or initiate a direct measure of a behaviour
- May be measured by something that occurs within the confines of the survey or something outside of the survey

Example 1: Active Information Choice

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`

Example 1: Active Information Choice

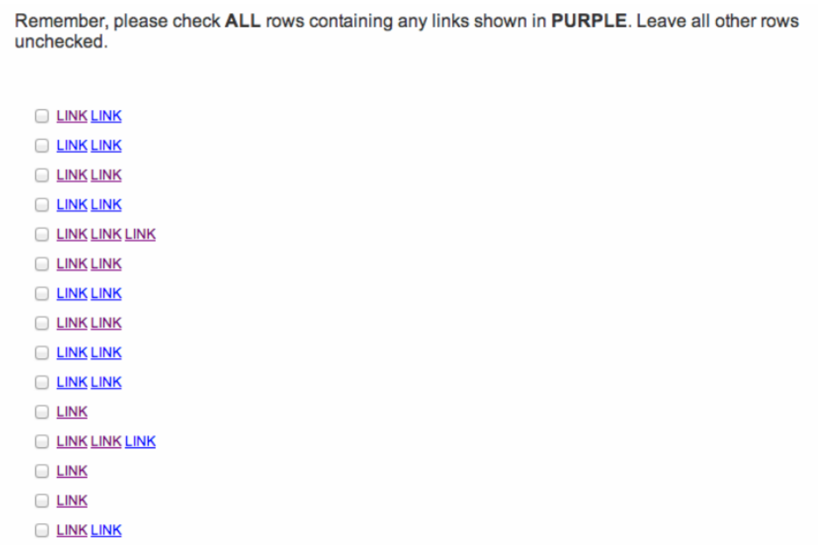
- "Followed link" identification[8]

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`



Remember, please check **ALL** rows containing any links shown in **PURPLE**. Leave all other rows unchecked.

- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK LINK
- [ ] LINK
- [ ] LINK LINK LINK
- [ ] LINK
- [ ] LINK
- [ ] LINK LINK

Example 1: Active Information Choice

- "Followed link" identification[8]

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`

Example 1: Active Information Choice

- "Followed link" identification[8]
- Information boards[9]

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`

Reports From the Hive, Where the Swarm Concurs	Doctors Can Work Together to Improve Patient Health, But Need Appropriate Incentives	SEC Vote Requires Business Filings to Add Environmental Risks to Bottom Line	Wellness, Rather Than Illness, Is Focus Under Outcome-Accountable Care
Pay for Performance Improves Quality of Health Care Through Collaborative Medicine	Patients Better Served When Providers Paid for Health Outcomes	Anatomy of a Tear-Jerker	Gender Differences in Education Need Innovative Solution
Why are 3-D Movies so Bad?	Improving America's Health Requires Provider Incentives, Not 'Fee-for-Service'	Spammers Use the Human Touch to Avoid CAPTCHA	Heart Attack While Dining at Heart Attack Grill in Las Vegas
Physicians Group Says Quality Will Improve Under Outcome-based Payments	When Paid for Outcomes, Doctors Have Little Reason to Treat Highest Risk Patients	USDA Raises Corn Export Outlook	Out of the O.R., T.R. Knight Back Onto the Stage
Council Is Set to Consider Increases in Hotel and Property Taxes	A Bowl of Chili with Bragging Rights	Will a Standardized System for Verifying Web Identity Ever Catch On?	Paying Doctors Based on Outcomes Will Lead to Rationing

Example 1: Active Information Choice

- "Followed link" identification[8]
- Information boards[9]

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`

Example 1: Active Information Choice

- "Followed link" identification[8]
- Information boards[9]
- Video choice[10]

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`

Example 1: Active Information Choice

- "Followed link" identification[8]
- Information boards[9]
- Video choice[10]
- Dynamic Process Tracing Environment [11]

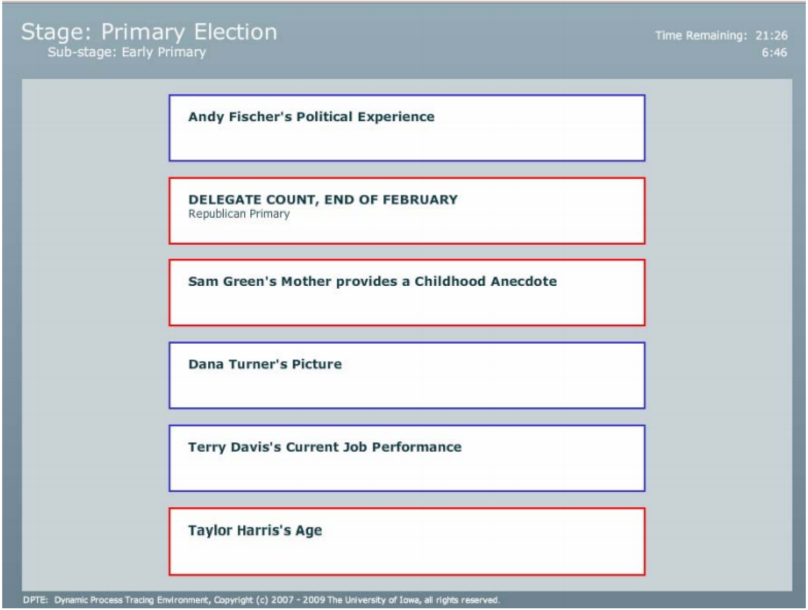
---

[8] Guess, AM. 2015. "Measure for Measure." *Political Analysis* 23: 59–75. doi:10.1093/pan/mpu010

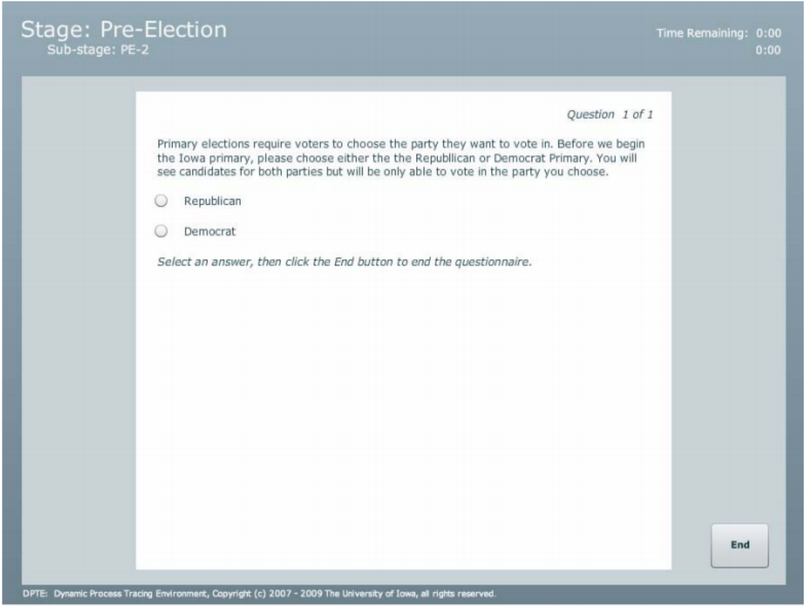
[9] Leeper, TJ. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46. doi:10.1093/poq/nft045

[10] Arceneaux, K & Johnson, M. 2012. *Changing Minds or Changign Channels*. Chicago: The University of Chicago Press.

[11] `https://dpte.polisci.uiowa.edu/dpte/`


Stage: Primary Election
Sub-stage: Early Primary
Time Remaining: 21:26
6:46
Andy Fischer's Political Experience
DELEGATE COUNT, END OF FEBRUARY
Republican Primary
Sam Green's Mother provides a Childhood Anecdote
Dana Turner's Picture
Terry Davis's Current Job Performance
Taylor Harris's Age
DPTE: Dynamic Process Tracing Environment, Copyright (c) 2007 - 2009 The University of Iowa, all rights reserved.


Iowa General Election
January, 2008
Time remaining: 5:23
Hillary Clinton wins in South Dakota!
DPTE: Dynamic Process Tracing Environment, Copyright (c) 2007 - 2008 The University of Iowa, all rights reserved.


Stage: Pre-Election
Sub-stage: PE-2
Time Remaining: 0:00
0:00
Question 1 of 1
Primary elections require voters to choose the party they want to vote in. Before we begin the Iowa primary, please choose either the the Republlican or Democrat Primary. You will see candidates for both parties but will be only able to vote in the party you choose.
Republican
Democrat
Select an answer, then click the End button to end the questionnaire.
End
DPTE: Dynamic Process Tracing Environment, Copyright (c) 2007 - 2009 The University of Iowa, all rights reserved.

Example 2: Sign-up/Enrolment

An extension of information choice behaviour would be explicit engagement in other kinds of (small) behaviours, such as:

- Entering an email address to receive information or join a mailing list [12] [13]
- Signing up for an appointment or further interaction

---

[12] Leeper, TJ. 2017. "How Does Treatment Self-Selection Affect Inferences About Political Communication?" *Journal of Experimental Political Science*: In press.

[13] Bolsen, Druckman, & Cook. 2014. "Communication and Collective Actions." *Journal of Experimental Political Science* 1(1): 24–38. doi:10.1017/xps.2014.2

Example 3: Incentivised Survey Questions

Definitions:

- A survey question is just a self-report
- An *incentivized* survey question attached financial gains or losses to the answer options

Mark your gamble selection with an **X** in the last column across from your preferred gamble.

Gamble	Event	Payoff	Probabilities	Your Selection
1	A	$10	50%	
	B	$10	50%	
2	A	$18	50%	
	B	$6	50%	
3	A	$26	50%	
	B	$2	50%	
4	A	$34	50%	
	B	-$2	50%	
5	A	$42	50%	
	B	-$6	50%	

Eckel & Grossman. 2008 "Forecasting risk attitudes." *Journal of Economic Behavior & Organization* 68(1): 1–17. doi:10.1016/j.jebo.2008.04.006

Example 3: Incentivised Survey Questions

Definitions:

- A survey question is just a self-report
- An *incentivized* survey question attached financial gains or losses to the answer options

Example 3: Incentivised Survey Questions

Definitions:

- A survey question is just a self-report
- An *incentivized* survey question attached financial gains or losses to the answer options

Paradigm could be applied to any measure of behavioural intentions to avoid cheap talk.

Example 4: Purchasing Decisions

Common ways to study purchasing behaviour include:

Example 4: Purchasing Decisions

Common ways to study purchasing behaviour include:

- Direct attitudinal questions

Example 4: Purchasing Decisions

Common ways to study purchasing behaviour include:

- Direct attitudinal questions
- Retrospective and prospective self-reports

Example 4: Purchasing Decisions

Common ways to study purchasing behaviour include:

- Direct attitudinal questions
- Retrospective and prospective self-reports
- Conjoint experiments

Example 4: Purchasing Decisions

Common ways to study purchasing behaviour include:

- Direct attitudinal questions
- Retrospective and prospective self-reports
- Conjoint experiments

Another way is embedding a purchase in a survey.[14]

[14] Bolsen, T. 2011. "A Lightbulb Goes On." *Political Behavior* 35(1): 1–20. 10.1007/s11109-011-9186-5

Source: Wikimedia Commons (Sun Ladder, KMJ)

Example 5: Donations

- Miller and Krosnick[15] asked for charitable donations via cheque directly as part of a paper-and-pencil survey

[15] Miller, Krosnick, & Lowe. N.d. "The Impact of Policy Change Threat on Financial Contributions to Interest Groups." Working paper.

[16] Klar & Piston. 2015. "The influence of competing organisational appeals on individual donations." *Journal of Public Policy* 35(2): 171–91. doi:10.1017/S0143814X15000203

Example 5: Donations

- Miller and Krosnick[15] asked for charitable donations via cheque directly as part of a paper-and-pencil survey
- Klar and Piston[16] offered respondents a survey incentive up-front for participation and then later offered them a chance to donate (a portion of payment) to a charity

---

[15] Miller, Krosnick, & Lowe. N.d. "The Impact of Policy Change Threat on Financial Contributions to Interest Groups." Working paper.

[16] Klar & Piston. 2015. "The influence of competing organisational appeals on individual donations." *Journal of Public Policy* 35(2): 171–91. doi:10.1017/S0143814X15000203

Example 6: Web Tracking Data

1. Active installation of a tracking app, such as YouGov Pulse[17] [18]
2. Post-hoc collection of web history files using something like Web Historian [19]

---

[17] https://yougov.co.uk/find-solutions/profiles/pulse/

[18] Guess, AM. N.d. "Media Choice and Moderation." Working paper, https://dl.dropboxusercontent.com/u/663930/GuessJMP.pdf.

[19] http://www.webhistorian.org/

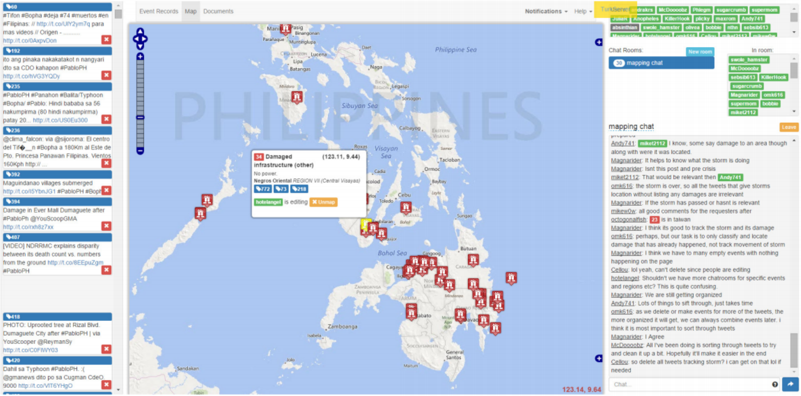
Other Possibilities

[20] Mao, Mason, Suri, Watts. 2016. "An Experimental Study of Team Size and Performance on a Complex Task." *PLoS ONE* 11(4): e0153048. doi:10.1371/journal.pone.0153048

Other Possibilities

- Coordination tasks
  - Synchronous group tasks[20]
  - Game play
  - Simulations

[20] Mao, Mason, Suri, Watts. 2016. "An Experimental Study of Team Size and Performance on a Complex Task." *PLoS ONE* 11(4): e0153048. doi:10.1371/journal.pone.0153048


Event Records
Map
Documents
Notifications
Help
PHILIPPINES
Philippine Sea
Sibuyan Sea
Visayan Sea
Bohol Sea
Damaged infrastructure (other)
(123.11, 9.44)
No power.
Negros Oriental REGION VII (Central Visayas)
hotelangel is editing
Unmap
123.14, 9.64
Chat Rooms:
New room
mapping chat
In room:
Leave
Chat...

Other Possibilities

- Coordination tasks
  - Synchronous group tasks[20]
  - Game play
  - Simulations

[20] Mao, Mason, Suri, Watts. 2016. "An Experimental Study of Team Size and Performance on a Complex Task." *PLoS ONE* 11(4): e0153048. doi:10.1371/journal.pone.0153048

Other Possibilities

- Coordination tasks
  - Synchronous group tasks[20]
  - Game play
  - Simulations
- Offering incentives to perform future behaviour (tracked elsewhere)

[20] Mao, Mason, Suri, Watts. 2016. "An Experimental Study of Team Size and Performance on a Complex Task." *PLoS ONE* 11(4): e0153048. doi:10.1371/journal.pone.0153048

Other Possibilities

- Coordination tasks
  - Synchronous group tasks[20]
  - Game play
  - Simulations
- Offering incentives to perform future behaviour (tracked elsewhere)
- OAuth/API integrations w/ other platforms
  - Merging website usage data w/ survey data
  - Treating website sign-up or usage as behavioural outcomes
  - Linking with smartphone metadata

---

[20] Mao, Mason, Suri, Watts. 2016. "An Experimental Study of Team Size and Performance on a Complex Task." *PLoS ONE* 11(4): e0153048. doi:10.1371/journal.pone.0153048

Some principles for survey measures of behaviour

Some principles for survey measures of behaviour

1. Know why you are collecting a behavioural measure!

Some principles for survey measures of behaviour

1. Know why you are collecting a behavioural measure!
2. Know whether you are studying a past, present, or future behaviour.

Some principles for survey measures of behaviour

1. Know why you are collecting a behavioural measure!
2. Know whether you are studying a past, present, or future behaviour.
3. Be creative! Recognise possibilities and limitations of any given survey mode.

Some principles for survey measures of behaviour

1. Know why you are collecting a behavioural measure!
2. Know whether you are studying a past, present, or future behaviour.
3. Be creative! Recognise possibilities and limitations of any given survey mode.
4. Validate, validate, validate!

Activity!

With a partner, brainstorm how one or more these behavioural measures might be applied to a survey experiment (either as outcome, treatment, covariate, or behavioural check) relevant to your own work or your organisation.