Lab Experiments
00000000000

Survey Experiments
0000000000000000

Generalizability
0000000

FLS 6441 - Methods III: Explanation and Causation

Week 4 - Survey and Lab Experiments

Jonathan Phillips

April 2019

Lab Experiments
00000000000

Survey Experiments
00000000000000000

Generalizability
0000000

Survey and Lab Experiments

▶ Why survey and lab experiments?

Lab Experiments
00000000000

Survey Experiments
000000000000000000

Generalizability
0000000

Survey and Lab Experiments

▶ Why survey and lab experiments?
  1. Treatments we cannot administer in reality

Survey and Lab Experiments

► Why survey and lab experiments?
   1. Treatments we cannot administer in reality
   2. Random treatment assignment not permitted in reality

Survey and Lab Experiments

- ▶ Why survey and lab experiments?
    1. Treatments we cannot administer in reality
    2. Random treatment assignment not permitted in reality
    3. Outcome measurements that are hard to take in reality

Survey and Lab Experiments

► Why survey and lab experiments?
  1. Treatments we cannot administer in reality
  2. Random treatment assignment not permitted in reality
  3. Outcome measurements that are hard to take in reality
  4. Reduce variation in context and noise in data

Survey and Lab Experiments

► Why survey and lab experiments?
  1. Treatments we cannot administer in reality
  2. Random treatment assignment not permitted in reality
  3. Outcome measurements that are hard to take in reality
  4. Reduce variation in context and noise in data
  5. To generalize beyond specific situations to abstract behaviour

Section 1

Lab Experiments

Lab Experiments

► **Treatment Assignment**: Same as a Field Experiment

Lab Experiments

▶ **Treatment Assignment**: Same as a Field Experiment
▶ **Treatment**: Not a manipulation of real world political or economic processes, but establishing controlled 'lab' conditions

Lab Experiments

► **Treatment Assignment**: Same as a Field Experiment
► **Treatment**: Not a manipulation of real world political or economic processes, but establishing controlled 'lab' conditions
   ► The advantage: Control over context helps isolate mechanisms

Lab Experiments

- **Treatment Assignment**: Same as a Field Experiment
- **Treatment**: Not a manipulation of real world political or economic processes, but establishing controlled 'lab' conditions
  - The advantage: Control over context helps isolate mechanisms
  - The disadvantage: Can we generalize to the real world?

Lab Experiments

► Problems generalizing from the lab:

Lab Experiments

- ▶ Problems generalizing from the lab:
  - ▶ **Hawthorne effect**: Lab context influences behaviour, social desirability bias

Lab Experiments

► Problems generalizing from the lab:
  ► **Hawthorne effect**: Lab context influences behaviour, social desirability bias
  ► **Context effects**: The real-world always provides more information, more history

Lab Experiments

- ▶ Problems generalizing from the lab:
    - ▶ **Hawthorne effect**: Lab context influences behaviour, social desirability bias
    - ▶ **Context effects**: The real-world always provides more information, more history
    - ▶ **Process effects**: People care *how* decisions are made
    - ▶ **Selection effects**: Actors in specific roles are rarely representative samples, 'WEIRD' or pro-social lab subjects

Lab Experiments

▶ The lab differs from the field

Lab Experiments

- ▶ The lab differs from the field
  - ▶ The stakes

Lab Experiments

- ▶ The lab differs from the field
  - ▶ The stakes
  - ▶ The norms (specific norms of being an experimental subject)

Lab Experiments

- ► The lab differs from the field
  - ► The stakes
  - ► The norms (specific norms of being an experimental subject)
  - ► The degree of scrutiny

Lab Experiments

- ▶ The lab differs from the field
  - ▶ The stakes
  - ▶ The norms (specific norms of being an experimental subject)
  - ▶ The degree of scrutiny
  - ▶ The sample of individuals

Lab Experiments

► The lab differs from the field
  ► The stakes
  ► The norms (specific norms of being an experimental subject)
  ► The degree of scrutiny
  ► The sample of individuals
  ► The degree of anonymity

Lab Experiments

▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)

Lab Experiments

- ▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)
- ▶ Decisions change depending on the degree of **scrutiny**

Lab Experiments

- ▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)
- ▶ Decisions change depending on the degree of **scrutiny**
  - ▶ "You tip more when you're on a date"

Lab Experiments

- ▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)
- ▶ Decisions change depending on the degree of **scrutiny**
    - ▶ "You tip more when you're on a date"
    - ▶ Social norms are activated, eg. treating one-shot games like repeated games

Lab Experiments   Survey Experiments   Generalizability

00000●00000   0000000000000000   0000000

Lab Experiments

- ▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)
- ▶ Decisions change depending on the degree of **scrutiny**
    - ▶ "You tip more when you're on a date"
    - ▶ Social norms are activated, eg. treating one-shot games like repeated games
    - ▶ Scrutiny alters who wants to make a decision as well as the decision they make

Lab Experiments

- ▶ Lab experiments are *inherently* imperfect (Levitt and List 2006)
- ▶ Decisions change depending on the degree of **scrutiny**
  - ▶ "You tip more when you're on a date"
  - ▶ Social norms are activated, eg. treating one-shot games like repeated games
  - ▶ Scrutiny alters who wants to make a decision as well as the decision they make
  - ▶ Subjects use cues (heuristics) to draw on 'similar' situations from the real world

Lab Experiments

▶ Many studies find more cooperation in the lab than in the real world

Lab Experiments

- ▶ Many studies find more cooperation in the lab than in the real world
  - ▶ Scrutiny increases cooperation

Lab Experiments

- ▶ Many studies find more cooperation in the lab than in the real world
  - ▶ Scrutiny increases cooperation
  - ▶ Anonymity reduces cooperation

Lab Experiments

- ▶ Many studies find more cooperation in the lab than in the real world
  - ▶ Scrutiny increases cooperation
  - ▶ Anonymity reduces cooperation
  - ▶ That's interesting in itself! We can manipulate the degree of scrutiny/anonymity etc.

Lab Experiments

- ► Many studies find more cooperation in the lab than in the real world
    - ► Scrutiny increases cooperation
    - ► Anonymity reduces cooperation
    - ► That's interesting in itself! We can manipulate the degree of scrutiny/anonymity etc.
- ► Lab experiments may be generalizable where norms/morality is less important (???)

Lab-in-the-Field Experiments

▶ In a natural setting with the target population

Lab-in-the-Field Experiments

- ▶ In a natural setting with the target population
- ▶ Standardized, artificial treatment and measurement

Lab-in-the-Field Experiments

► Habyarimana et al (2007)

Lab-in-the-Field Experiments

- ▶ Habyarimana et al (2007)
- ▶ Existing consensus: Ethnic diversity -> **Less** public goods provision

Lab-in-the-Field Experiments

- ▶ Habyarimana et al (2007)
- ▶ Existing consensus: Ethnic diversity -> **Less** public goods provision
- ▶ But how? Theories:

Lab-in-the-Field Experiments

- ▶ Habyarimana et al (2007)
- ▶ Existing consensus: Ethnic diversity -> **Less** public goods provision
- ▶ But how? Theories:
  - ▶ Preferences - in-group fairness
  - ▶ Technology - social networks permit identification and sanctioning
  - ▶ Strategy Selection - choose to cooperate more often

Lab-in-the-Field Experiments

- ▶ Lab-in-the-field
- ▶ **Population**: Ugandans
- ▶ **Sample**: 300 people in a diverse area with few public goods
- ▶ **Treatment/Control**: Various Games
- ▶ **Treatment assignment**: Random assignment to co-ethnic/non-co-ethnic

Lab-in-the-Field Experiments

▶ **Preferences** - dictator game between self and two others
  ▶ No bias towards co-ethnics

Lab-in-the-Field Experiments

- **Preferences** - dictator game between self and two others
  - No bias towards co-ethnics
- **Technology 1, productivity** - teamwork in a puzzle requiring communication
  - Co-ethnic teams don't perform any better

Lab-in-the-Field Experiments

- **Preferences** - dictator game between self and two others
  - No bias towards co-ethnics
- **Technology 1, productivity** - teamwork in a puzzle requiring communication
  - Co-ethnic teams don't perform any better
- **Technology 2, social networks** - Can you find a co-ethnic in the town faster than a non-co-ethnic?
  - Yes (43% vs 28% success)

Lab-in-the-Field Experiments

- **Preferences** - dictator game between self and two others
    - No bias towards co-ethnics
- **Technology 1, productivity** - teamwork in a puzzle requiring communication
    - Co-ethnic teams don't perform any better
- **Technology 2, social networks** - Can you find a co-ethnic in the town faster than a non-co-ethnic?
    - Yes (43% vs 28% success)
- **Strategy Selection** - Does anonymity for the sender in the dictator game make a difference?
    - Yes - offer more to co-ethnics when offerers believe they can be seen

Lab-in-the-Field Experiments

- **Preferences** - dictator game between self and two others
    - No bias towards co-ethnics
- **Technology 1, productivity** - teamwork in a puzzle requiring communication
    - Co-ethnic teams don't perform any better
- **Technology 2, social networks** - Can you find a co-ethnic in the town faster than a non-co-ethnic?
    - Yes (43% vs 28% success)
- **Strategy Selection** - Does anonymity for the sender in the dictator game make a difference?
    - Yes - offer more to co-ethnics when offerers believe they can be seen

Lab-in-the-Field Experiments

▶ **Conclusion:** Norms and Networks allow co-ethnics to provide more public goods

Lab-in-the-Field Experiments

- **Conclusion:** Norms and Networks allow co-ethnics to provide more public goods
  - ...But where are the public goods here?
  - Are public goods organized by voluntary contributions or coercive central authority?
  - Is this true of all parts of Kampala? Uganda? All ethnic groups?

Section 2

Survey Experiments

Survey Experiments

▶ Treatment occurs *within* the survey questionnaire

Survey Experiments

▶ Treatment occurs *within* the survey questionnaire
▶ Outcome measurement also *within* the survey questionnaire

Survey Experiments

- ▶ Treatment occurs *within* the survey questionnaire
- ▶ Outcome measurement also *within* the survey questionnaire

  - ▶ Different versions of the questionnaire randomly applied

Survey Experiments

- ▶ Treatment occurs *within* the survey questionnaire
- ▶ Outcome measurement also *within* the survey questionnaire

  - ▶ Different versions of the questionnaire randomly applied
  - ▶ Not a field experiment: Still an artificial context

Survey Experiments

► Treatment occurs *within* the survey questionnaire
► Outcome measurement also *within* the survey questionnaire

   ► Different versions of the questionnaire randomly applied
   ► Not a field experiment: Still an artificial context
   ► Not a lab experiment: People not brought to a single location
      or interacting

Lab Experiments
00000000000

Survey Experiments
000●000000000000000

Generalizability
0000000

Survey Experiments

▶ Easy and cheap to implement

Survey Experiments

- ▶ Easy and cheap to implement
- ▶ Can be targeted to our real population of interest

Survey Experiments

- ▶ Easy and cheap to implement
- ▶ Can be targeted to our real population of interest
- ▶ But a limited range of 'weak' treatments possible

Survey Experiments

- ▶ Easy and cheap to implement
- ▶ Can be targeted to our real population of interest
- ▶ But a limited range of 'weak' treatments possible
- ▶ And we can only measure short-term effects

Survey Experiments

▶ Humans are subject to psychological and social influences

Survey Experiments

- ▶ Humans are subject to psychological and social influences
- ▶ These create threats to estimating causal effects

Survey Experiments

► Humans are subject to psychological and social influences
► These create threats to estimating causal effects
   ► **Social Desirability Bias:** Respondents lie when they think someone is listening to their answers! (Including the enumerator)

Survey Experiments

► Humans are subject to psychological and social influences
► These create threats to estimating causal effects
  ► **Social Desirability Bias:** Respondents lie when they think someone is listening to their answers! (Including the enumerator)
  ► **Sequencing Bias:** If we ask about who you voted for after twenty questions about redistribution and equality, your answer might be different

Survey Experiments

▶ Humans are subject to psychological and social influences
▶ These create threats to estimating causal effects
  ▶ **Social Desirability Bias:** Respondents lie when they think someone is listening to their answers! (Including the enumerator)
  ▶ **Sequencing Bias:** If we ask about who you voted for after twenty questions about redistribution and equality, your answer might be different
  ▶ **Acquiescence Bias:** Thinking about your answers is hard, so it's easier just to agree with the default/first option

Survey Experiments

▶ Humans are subject to psychological and social influences
▶ These create threats to estimating causal effects
  ▶ **Social Desirability Bias:** Respondents lie when they think someone is listening to their answers! (Including the enumerator)
  ▶ **Sequencing Bias:** If we ask about who you voted for after twenty questions about redistribution and equality, your answer might be different
  ▶ **Acquiescence Bias:** Thinking about your answers is hard, so it's easier just to agree with the default/first option
  ▶ **Anchoring Bias:** The first piece of information in a question affects our response, Eg. The average person does x, what do you do?

Types of Survey Experiments

▶ But we can also use these influences to our advantage to study psychological and social processes:

Types of Survey Experiments

- ▶ But we can also use these influences to our advantage to study psychological and social processes:

1. **Framing Experiments** - how responses vary to question content

Types of Survey Experiments

- ▶ But we can also use these influences to our advantage to study psychological and social processes:
1. **Framing Experiments** - how responses vary to question content
2. **Priming Experiments** - to measure the effect of a prime on a response to a fixed question

Types of Survey Experiments

▶ But we can also use these influences to our advantage to
study psychological and social processes:

1. **Framing Experiments** - how responses vary to question
content

2. **Priming Experiments** - to measure the effect of a prime
on a response to a fixed question

3. **List Experiments** - to reduce social desirability bias in
measurement

Types of Survey Experiments

▶ But we can also use these influences to our advantage to study psychological and social processes:

1. **Framing Experiments** - how responses vary to question content
2. **Priming Experiments** - to measure the effect of a prime on a response to a fixed question
3. **List Experiments** - to reduce social desirability bias in measurement
4. **Conjoint Experiments** - to measure relative preferences

1. Framing

▶ How much do details in the question ('wording') affect our responses?

1. Framing

- ▶ How much do details in the question ('wording') affect our responses?
- ▶ (i) Town A has only **80%** of the income of Town B, and the gap is widening. The government proposes to transfer income from Town A to Town B to reduce inequality. How much do you think would be a fair tax on Town A's income?

1. Framing

► How much do details in the question ('wording') affect our responses?

► (i) Town A has only **80%** of the income of Town B, and the gap is widening. The government proposes to transfer income from Town A to Town B to reduce inequality. How much do you think would be a fair tax on Town A's income?

► (ii) Town A has only **20%** of the income of Town B, and the gap is widening. The government proposes to transfer income from Town A to Town B to reduce inequality. How much do you think would be a fair tax on Town A's income?

1. Framing

▶ How much do details in the question ('wording') affect our responses?

▶ (i) Town A has only **80%** of the income of Town B, and the gap is widening. The government proposes to transfer income from Town A to Town B to reduce inequality. How much do you think would be a fair tax on Town A's income?

▶ (ii) Town A has only **20%** of the income of Town B, and the gap is widening. The government proposes to transfer income from Town A to Town B to reduce inequality. How much do you think would be a fair tax on Town A's income?

▶ 1%

▶ 5%

▶ 10%

▶ 25%

▶ 50%

1. Framing

▶ Within/Between Survey Experiments

1. Framing

- ▶ Within/Between Survey Experiments
- ▶ Between: Treated and Control are different people

1. Framing

- ▶ Within/Between Survey Experiments
- ▶ Between: Treated and Control are different people
- ▶ Within: Treated and Control measures from the same person

  - ▶ But aren't these different 'units'??

1. Framing

- ▶ Within/Between Survey Experiments
- ▶ Between: Treated and Control are different people
- ▶ Within: Treated and Control measures from the same person

    - ▶ But aren't these different 'units'?? Yes!

1. Framing

- ▶ Within/Between Survey Experiments
- ▶ Between: Treated and Control are different people
- ▶ Within: Treated and Control measures from the same person
    - ▶ But aren't these different 'units'?? Yes!
    - ▶ But the time difference is usually just a few minutes, so maybe more plausible

1. Framing

- ▶ Within/Between Survey Experiments
- ▶ Between: Treated and Control are different people
- ▶ Within: Treated and Control measures from the same person
    - ▶ But aren't these different 'units'?? Yes!
    - ▶ But the time difference is usually just a few minutes, so maybe more plausible
    - ▶ More problematic is sequencing bias

1. Framing

▶ Within/Between Survey Experiments
▶ Between: Treated and Control are different people
▶ Within: Treated and Control measures from the same person

     ▶ But aren't these different 'units'?? Yes!
     ▶ But the time difference is usually just a few minutes, so maybe more plausible
     ▶ More problematic is sequencing bias
     ▶ But we can also randomize the sequence

2. Priming

▶ The entire point of survey bias is that the questions we ask
  change the answers we get

2. Priming

- ▶ The entire point of survey bias is that the questions we ask change the answers we get
- ▶ So changing questions across respondents might change *ALL* subsequent responses

2. Priming

- ▶ The entire point of survey bias is that the questions we ask change the answers we get
- ▶ So changing questions across respondents might change *ALL* subsequent responses
- ▶ We design a prior task that creates an *unconscious* bias in subsequent answers

2. Priming

- ▶ The entire point of survey bias is that the questions we ask change the answers we get
- ▶ So changing questions across respondents might change *ALL* subsequent responses
- ▶ We design a prior task that creates an *unconscious* bias in subsequent answers
- ▶ The treatment/control prime is separated from the response question

2. Priming

- ► The entire point of survey bias is that the questions we ask change the answers we get
- ► So changing questions across respondents might change *ALL* subsequent responses
- ► We design a prior task that creates an *unconscious* bias in subsequent answers
- ► The treatment/control prime is separated from the response question
  - ► Usually the prime is a few questions before the response

2. Priming

- ▶ The entire point of survey bias is that the questions we ask change the answers we get
- ▶ So changing questions across respondents might change *ALL* subsequent responses
- ▶ We design a prior task that creates an *unconscious* bias in subsequent answers
- ▶ The treatment/control prime is separated from the response question
    - ▶ Usually the prime is a few questions before the response
    - ▶ Everyone sees the same response question

2. Priming

► Eg. If we want to measure the effect of 'hope' vs. 'fear' on
   the demand for constitutional reform:

2. Priming

- ▶ Eg. If we want to measure the effect of 'hope' vs. 'fear' on the demand for constitutional reform:
- ▶ There are good reasons to be hopeful for democracy: More people voted than ever before in the 2018 election, suggesting strong citizen participation in democracy. Did you vote in 2018?

2. Priming

▶ Eg. If we want to measure the effect of 'hope' vs. 'fear' on the demand for constitutional reform:

▶ There are good reasons to be hopeful for democracy: More people voted than ever before in the 2018 election, suggesting strong citizen participation in democracy. Did you vote in 2018?

▶ Many people fear for the survival of democracy: The proportion of voters who turned out to vote fell again in the 2018 election, suggesting a critical problem of declining trust and engagement in democracy. Did you vote in 2018?

2. Priming

▶ Eg. If we want to measure the effect of 'hope' vs. 'fear' on the demand for constitutional reform:

▶ There are good reasons to be hopeful for democracy: More people voted than ever before in the 2018 election, suggesting strong citizen participation in democracy. Did you vote in 2018?

▶ Many people fear for the survival of democracy: The proportion of voters who turned out to vote fell again in the 2018 election, suggesting a critical problem of declining trust and engagement in democracy. Did you vote in 2018?

▶ How much do you support constitutional reform?

3. List Experiments

► Most survey responses are biased to give the answer they think the researcher wants

3. List Experiments

► Most survey responses are biased to give the answer they
   think the researcher wants
   ► **Social desirability bias** has *differential* effects across
     respondents and topics
     ► Most people say they recycle, even though they do not

3. List Experiments

- ▶ Most survey responses are biased to give the answer they think the researcher wants
  - ▶ **Social desirability bias** has *differential* effects across respondents and topics
    - ▶ Most people say they recycle, even though they do not
    - ▶ But rich people exaggerate more than poor people

3. List Experiments

- ▶ Most survey responses are biased to give the answer they think the researcher wants
  - ▶ **Social desirability bias** has *differential* effects across respondents and topics
    - ▶ Most people say they recycle, even though they do not
    - ▶ But rich people exaggerate more than poor people
- ▶ List experiments make individual responses *invisible* to the researcher

3. List Experiments

- ▶ Most survey responses are biased to give the answer they think the researcher wants
  - ▶ **Social desirability bias** has *differential* effects across respondents and topics
    - ▶ Most people say they recycle, even though they do not
    - ▶ But rich people exaggerate more than poor people
- ▶ List experiments make individual responses *invisible* to the researcher
  - ▶ Knowing this, hopefully the respondent answers more accurately

3. List Experiments

I am now going to read out a list of activities. Please count the number of these activities that you have done in the past one year. Please do not tell me WHICH activities you have done, only the TOTAL NUMBER of them:

- ▶ Voted
- ▶ Attended a Town Hall Meeting
- ▶ Travelled to the State Capital

3. List Experiments

I am now going to read out a list of activities. Please count the number of these activities that you have done in the past one year. Please do not tell me WHICH activities you have done, only the TOTAL NUMBER of them:

- ► Voted
- ► Attended a Town Hall Meeting
- ► Been offered a gift, some food or money in exchange for your vote;
- ► Travelled to the State Capital

3. List Experiments

►

► Eg. Gonzalez-Ocantos (2010) - list experiment on vote-buying in Nicaragua 2008 municipal elections

► **Direct Question**: Have you received a gift or favour in exchange for your vote?

     ► 3%

3. List Experiments

- ▶
- ▶ Eg. Gonzalez-Ocantos (2010) - list experiment on vote-buying in Nicaragua 2008 municipal elections
- ▶ **Direct Question**: Have you received a gift or favour in exchange for your vote?
    - ▶ 3%
- ▶ **List experiment**:
    - ▶ Just the difference in mean responses between treatment and control lists
    - ▶ 24% = 2.31 - 2.06

3. List Experiments

Assumptions:

1. No Liars - People answer honestly in the presence of the sensitive item

3. List Experiments

Assumptions:

1. No Liars - People answer honestly in the presence of the sensitive item
   - ▶ Do respondents really understand anonymity?

3. List Experiments

Assumptions:

1. No Liars - People answer honestly in the presence of the sensitive item
   - ▶ Do respondents really understand anonymity?
2. No Ceiling effects - '4' means my answers are no longer anonymous; instead report '3'

3. List Experiments

Assumptions:

1. No Liars - People answer honestly in the presence of the sensitive item
   - ▶ Do respondents really understand anonymity?
2. No Ceiling effects - '4' means my answers are no longer anonymous; instead report '3'
3. No Floor Effects - If the control items are rare, respondents may be reluctant to report '1' and choose '0' instead.

3. List Experiments

Assumptions:

1. No Liars - People answer honestly in the presence of the sensitive item
   - ▶ Do respondents really understand anonymity?
2. No Ceiling effects - '4' means my answers are no longer anonymous; instead report '3'
3. No Floor Effects - If the control items are rare, respondents may be reluctant to report '1' and choose '0' instead.
4. No Design Effects- Presence of the treatment item doesn't affect answers on other items
   - ▶ Bias towards a 'reasonable'/central number?

4. Conjoint Survey Experiments

▶ How do people make choices between many options?

4. Conjoint Survey Experiments

- ▶ How do people make choices between many options?
- ▶ Treatments are often 'bundles' of characteristics, but which aspect matters most?

4. Conjoint Survey Experiments

- ► How do people make choices between many options?
- ► Treatments are often 'bundles' of characteristics, but which aspect matters most?
- ► Also a problem of social desirability bias in which characteristics matter

4. Conjoint Survey Experiments

► Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?

4. Conjoint Survey Experiments

▶ Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?

▶ Vary education, profession, language, gender, national origin, etc.

4. Conjoint Survey Experiments

- ▶ Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- ▶ Vary education, profession, language, gender, national origin, etc.
- ▶ Profiles
  - ▶ Attributes
    - ▶ Values

4. Conjoint Survey Experiments

- ▶ Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- ▶ Vary education, profession, language, gender, national origin, etc.
- ▶ Profiles
  - ▶ Attributes
    - ▶ Values
- ▶ Randomize values and attribute order to prevent bias

4. Conjoint Survey Experiments

- ▶ Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- ▶ Vary education, profession, language, gender, national origin, etc.
- ▶ Profiles
  - ▶ Attributes
    - ▶ Values
- ▶ Randomize values and attribute order to prevent bias
- ▶ Treatment is the **combination** of attributes the respondent sees

4. Conjoint Survey Experiments

▶ Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
▶ Vary education, profession, language, gender, national origin, etc.
▶ Profiles
  ▶ Attributes
    ▶ Values
▶ Randomize values and attribute order to prevent bias
▶ Treatment is the **combination** of attributes the respondent sees
▶ Millions of possible treatments

Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	○	○

On a scale from 1 to 7, where 1 indicates that the United States should absolutely not admit the immigrant and 7 indicates that the United States should definitely admit the immigrant, how would you rate Immigrant 1?

Absolutely Not Admit						Definitely Admit
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Using the same scale, how would you rate Immigrant 2?

Absolutely Not Admit						Definitely Admit
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Fig. 1 Experimental design: Immigration conjoint. This figure illustrates the experimental design for the conjoint analysis that examines immigrant admission to the United States.

choice outcomes hereafter. Second, in "rating-based conjoint analysis," respondents give a numerical rating to each profile which represents their degree of preference for the profile. This format is preferred by some analysts who contend that such ratings provide more direct, finely grained information about respondents' preferences. We call this latter type of outcome a *rating outcome*.
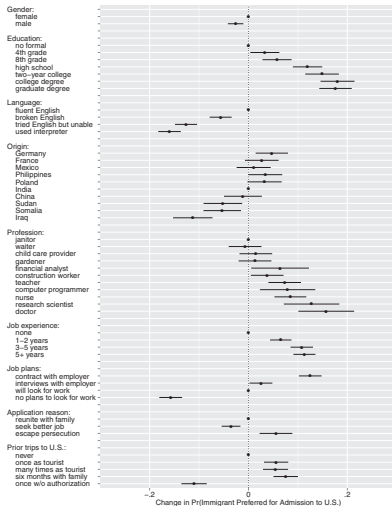
Fig. 3 Effects of immigrant attributes on preference for admission. This plot shows estimates of the effects of the randomly assigned immigrant attributes on the probability of being preferred for admission to the United States. Estimates are based on the regression estimators with clustered standard errors; bars represent 95% confidence intervals. The points without horizontal bars denote the attribute value that is the reference category for each attribute.

4. Conjoint Survey Experiments

► Estimating results uses a simple regression of respondent choices on profile attribute-values

4. Conjoint Survey Experiments

▶ Estimating results uses a simple regression of respondent choices on profile attribute-values

▶ But each specific profile (treatment) may arise too rarely to make comparisons of individual attribute-values

4. Conjoint Survey Experiments

- ▶ Estimating results uses a simple regression of respondent choices on profile attribute-values
- ▶ But each specific profile (treatment) may arise too rarely to make comparisons of individual attribute-values
  - ▶ So this is **not** an Average Treatment Effect for each profile

4. Conjoint Survey Experiments

- ▶ Estimating results uses a simple regression of respondent choices on profile attribute-values
- ▶ But each specific profile (treatment) may arise too rarely to make comparisons of individual attribute-values
  - ▶ So this is **not** an Average Treatment Effect for each profile
  - ▶ Eg. the effect of gender when age, language etc. are held constant

4. Conjoint Survey Experiments

► Estimating results uses a simple regression of respondent choices on profile attribute-values

► But each specific profile (treatment) may arise too rarely to make comparisons of individual attribute-values

   ► So this is **not** an Average Treatment Effect for each profile

   ► Eg. the effect of gender when age, language etc. are held constant

   ► It is an **Average Marginal Component Effect**

4. Conjoint Survey Experiments

- ▶ Estimating results uses a simple regression of respondent choices on profile attribute-values
- ▶ But each specific profile (treatment) may arise too rarely to make comparisons of individual attribute-values
    - ▶ So this is **not** an Average Treatment Effect for each profile
    - ▶ Eg. the effect of gender when age, language etc. are held constant
    - ▶ It is an **Average Marginal Component Effect**
    - ▶ Eg. the effect of gender averaging across all possibilities of age, language, etc.

4. Conjoint Survey Experiments

Assumptions:

▶ We're still assuming people try to answer honestly

4. Conjoint Survey Experiments

Assumptions:

► We're still assuming people try to answer honestly

► The ordering of attributes does not matter (or is randomized)

4. Conjoint Survey Experiments

Assumptions:

- ▶ We're still assuming people try to answer honestly
- ▶ The ordering of attributes does not matter (or is randomized)
- ▶ Profiles are randomized

Section 3

Generalizability

Generalizability

▶ Can we generalize from survey/lab responses to real-world behaviour?

Generalizability

- ▶ Can we generalize from survey/lab responses to real-world behaviour?
- ▶ **Non-Behavioural Measures:**
  - ▶ What is at stake in the answer? Are there any actual consequences?

Generalizability

- ► Can we generalize from survey/lab responses to real-world behaviour?
- ► **Non-Behavioural Measures:**
  - ► What is at stake in the answer? Are there any actual consequences?
  - ► Will they have to defend their answer in the community later?

Generalizability

▶ Can we generalize from survey/lab responses to real-world behaviour?

▶ **Non-Behavioural Measures:**
   ▶ What is at stake in the answer? Are there any actual consequences?
   ▶ Will they have to defend their answer in the community later?
   ▶ Cognitive costs of thinking about your response

Generalizability

- ► Can we generalize from survey/lab responses to real-world behaviour?
- ► **Non-Behavioural Measures:**
    - ► What is at stake in the answer? Are there any actual consequences?
    - ► Will they have to defend their answer in the community later?
    - ► Cognitive costs of thinking about your response
    - ► 'Cheap talk'

▶ Can we generalize from survey/lab responses to real-world behaviour?

Can we generalize from survey/lab responses to real-world behaviour?

Credibility:

Can we generalize from survey/lab responses to real-world behaviour?

Credibility:

'Treatments' in survey experiments are just information or wording

Can we generalize from survey/lab responses to real-world behaviour?

Credibility:

'Treatments' in survey experiments are just information or wording

But do respondents 'believe' that information?

Can we generalize from survey/lab responses to real-world behaviour?

Credibility:

'Treatments' in survey experiments are just information or wording

But do respondents 'believe' that information?

Do they have conflicting information? What is their 'prior'?

Can we generalize from survey/lab responses to real-world behaviour?

Credibility:

'Treatments' in survey experiments are just information or wording

But do respondents 'believe' that information?

Do they have conflicting information? What is their 'prior'?

What 'authority' or 'trust' does the source (you!) have?

Can we generalize from survey/lab responses to real-world behaviour?

Can we generalize from survey/lab responses to real-world behaviour?

Context:

Can we generalize from survey/lab responses to real-world behaviour?

Context:

Our interpretation of treatments depends on subtle signals - someone telling you a Trump voter is moving in next door is very different to actually meeting that person

Can we generalize from survey/lab responses to real-world behaviour?

Context:

Our interpretation of treatments depends on subtle signals - someone telling you a Trump voter is moving in next door is very different to actually meeting that person

We want to abstract from that complexity, but are humans capable of reporting their 'average' responses?

Can we generalize from survey/lab responses to real-world behaviour?

Context:

Our interpretation of treatments depends on subtle signals - someone telling you a Trump voter is moving in next door is very different to actually meeting that person

We want to abstract from that complexity, but are humans capable of reporting their 'average' responses?

Careful planning of question sequencing (and randomization of question order)

Can we generalize from survey/lab responses to real-world behaviour?

Can we generalize from survey/lab responses to real-world behaviour?

Durability:

Can we generalize from survey/lab responses to real-world behaviour?

Durability:

We find that a nationalism prompt produces pro-statist attitudes five minutes later in a survey

Can we generalize from survey/lab responses to real-world behaviour?

Durability:

We find that a nationalism prompt produces pro-statist attitudes five minutes later in a survey

Would that effect persist one hour later?

Can we generalize from survey/lab responses to real-world
behaviour?

Durability:

We find that a nationalism prompt produces pro-statist attitudes five
minutes later in a survey

Would that effect persist one hour later?

How about a year later?

Can we generalize from survey/lab responses to real-world behaviour?

Durability:

We find that a nationalism prompt produces pro-statist attitudes five minutes later in a survey

Would that effect persist one hour later?

How about a year later?

How much has the respondent been exposed to these treatments previously? To competing treatments? Are there diminishing or accumulated effects?

Can we generalize from survey/lab responses to real-world behaviour?

Durability:

We find that a nationalism prompt produces pro-statist attitudes five minutes later in a survey

Would that effect persist one hour later?

How about a year later?

How much has the respondent been exposed to these treatments previously? To competing treatments? Are there diminishing or accumulated effects?

Real-world treatments are often continuous or repeated. We need to compare with when, where, and how real-world treatments happen.

Lab Experiments
00000000000

Survey Experiments
000000000000000000

Generalizability
000000●0

Generalizability

How reliable are the responses to a Conjoint Experiment?

Generalizability

How reliable are the responses to a Conjoint Experiment?

Stated preferences vs. Revealed preferences

Generalizability

How reliable are the responses to a Conjoint Experiment?

Stated preferences vs. Revealed preferences

Hainmueller et al 2014 - compare conjoint responses to a Swiss referendum

Generalizability

How reliable are the responses to a Conjoint Experiment?

Stated preferences vs. Revealed preferences

Hainmueller et al 2014 - compare conjoint responses to a Swiss referendum

Citizens voted on specific naturalization applicants (Really!)

Figure S11: Effects of Applicant Attributes on Opposition to Naturalization Request (Unweighted Survey Sample)
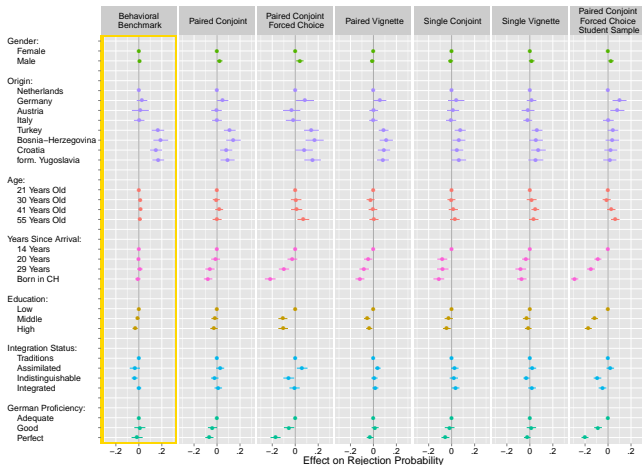

Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Generalizability

But note the conjoint method still hugely under-estimated the
overall rejection rate

21% versus 37% in reality