FLS 6441 - Methods III: Explanation and Causation

Week 4 - Survey and Lab Experiments

Jonathan Phillips

April 2019

Survey and Lab Experiments

- Why survey and lab experiments?

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

- 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
  - Subjets 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
  - Different versions of the questionnaire randomly applied

Survey Experiments

- Treatment occurs *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
  - 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

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 outcome measurement normally takes place immediately

Types of Survey Experiments

1. Framing Experiments

Types of Survey Experiments

1. Framing Experiments
2. Priming Experiments

Types of Survey Experiments

1. Framing Experiments
2. Priming Experiments
3. List Experiments

Types of Survey Experiments

1. Framing Experiments
2. Priming Experiments
3. List Experiments
4. Conjoint Experiments

Survey Experiments

- People responded differently to being told 'A' instead of 'B'. How do we interpret this?

Survey Experiments

- People responded differently to being told 'A' instead of 'B'. How do we interpret this?
  1. They were told 'A'/'B' by a survey enumerator - do they trust them? What is the source?

Survey Experiments

- People responded differently to being told 'A' instead of 'B'. How do we interpret this?
  1. They were told 'A'/'B' by a survey enumerator - do they trust them? What is the source?
  2. Are 'A'/'B' things that they would hear in the real world? In what context?

Survey Experiments

- People responded differently to being told 'A' instead of 'B'. How do we interpret this?
  1. They were told 'A'/'B' by a survey enumerator - do they trust them? What is the source?
  2. Are 'A'/'B' things that they would hear in the real world? In what context?
  3. What are they communicating in their answer? To impress the surveyor? Who is listening to their answers?

Survey Experiments

- People responded differently to being told 'A' instead of 'B'. How do we interpret this?
  1. They were told 'A'/'B' by a survey enumerator - do they trust them? What is the source?
  2. Are 'A'/'B' things that they would hear in the real world? In what context?
  3. What are they communicating in their answer? To impress the surveyor? Who is listening to their answers?
  4. What is at stake in the answer? Are there any actual consequences? Will they have to defend their answer in the community later? 'Cheap talk'

Section 3

Psychological Considerations

Psychology of Surveys

- People's responses are influenced by subtle features of the context

Psychology of Surveys

- People's responses are influenced by subtle features of the context
- These are a threat to identifying causal effects

Psychology of Surveys

- People's responses are influenced by subtle features of the context
- These are a threat to identifying causal effects
- But we can also use these biases to our advantage

1. Priming
2. Framing
3. Anchoring
4. Contamination

1. Priming

- A prior task that creates an unconscious bias in subsequent answers

1. Priming

- A prior task that creates an unconscious bias in subsequent answers
- Eg. We remind half of respondents about national Independence Day
- Then ask what they think about immigration
- Allowing us to measure the effect of 'nationalism' on migration attitudes

2. Framing

- How much do details in the question affect our responses?

2. Framing

- How much do details in the question affect our responses?
- Eg. A female citizen goes to her representative for help. How likely is she to receive help?

2. Framing

- How much do details in the question affect our responses?
- Eg. A female citizen goes to her representative for help. How likely is she to receive help?
- Eg. A male citizen goes to her representative for help. How likely is she to receive help?

4. Contamination

4. Contamination

- The entire point of survey experiments is that the questions we ask change the answers we get

4. Contamination

- The entire point of survey experiments is that the questions we ask change the answers we get
- So the answers to every question depend on the previous questions

4. Contamination

- The entire point of survey experiments is that the questions we ask change the answers we get
- So the answers to every question depend on the previous questions
- Usually affects all respondents equally

4. Contamination

- The entire point of survey experiments is that the questions we ask change the answers we get
- So the answers to every question depend on the previous questions
- Usually affects all respondents equally
- But survey experiments that vary across respondents might change *ALL* subsequent responses

Survey Experiments

- Within/Between Survey Experiments

Survey Experiments

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

Survey Experiments

- 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'??

Survey Experiments

- 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!

Survey Experiments

- 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

Survey Experiments

- 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 contamination

Survey Experiments

1. Vignettes -
2. List Experiments - reduce social desirability bias
3. Conjoint Experiments - measure preferences

List Experiments

- Survey experiments are valuable for measurement

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher
  - **Social desirability bias** has differential effects across respondents and topics
    - Most people say they recycle, even though they do not

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher
  - **Social desirability bias** has differential effects across respondents and topics
    - Most people say they recycle, even though they do not
    - Rich people lie more than poor people

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher
  - **Social desirability bias** has differential effects across respondents and topics
    - Most people say they recycle, even though they do not
    - Rich people lie more than poor people
- List experiments make individual responses *invisible* to the researcher

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher
  - **Social desirability bias** has differential effects across respondents and topics
    - Most people say they recycle, even though they do not
    - Rich people lie more than poor people
- List experiments make individual responses *invisible* to the researcher
- Knowing this, hopefully the respondent answers more accurately

List Experiments

- Survey experiments are valuable for measurement
  - Most survey responses are biased to impress the researcher
  - **Social desirability bias** has differential effects across respondents and topics
    - Most people say they recycle, even though they do not
    - Rich people lie more than poor people
- List experiments make individual responses *invisible* to the researcher
- Knowing this, hopefully the respondent answers more accurately
- Gonzalez-Ocantos et al (2010) - list experiment on vote-buying

List Experiments

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

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

List Experiments

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

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

List Experiments

- Nicaragua 2008 municipal elections
- **Direct Question**: Have you received a gift or favour in exchange for your vote?
  - 3%

List Experiments

- 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

List Experiments

- Do respondents really understand anonymity?

List Experiments

- Do respondents really understand anonymity?
- Floor/ceiling effects - 0 or 5 mean my answers are no longer anonymous

List Experiments

- Do respondents really understand anonymity?
- Floor/ceiling effects - 0 or 5 mean my answers are no longer anonymous
- Design Effects- Presence of treatment item doesn't affect answers on other items

List Experiments

- Do respondents really understand anonymity?
- Floor/ceiling effects - 0 or 5 mean my answers are no longer anonymous
- Design Effects- Presence of treatment item doesn't affect answers on other items
- Bias towards a 'reasonable'/central number?

Conjoint Survey Experiments

- How do people make choices between many options?

Conjoint Survey Experiments

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

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

Conjoint Survey Experiments

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

Conjoint Survey Experiments

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

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

Conjoint Survey Experiments

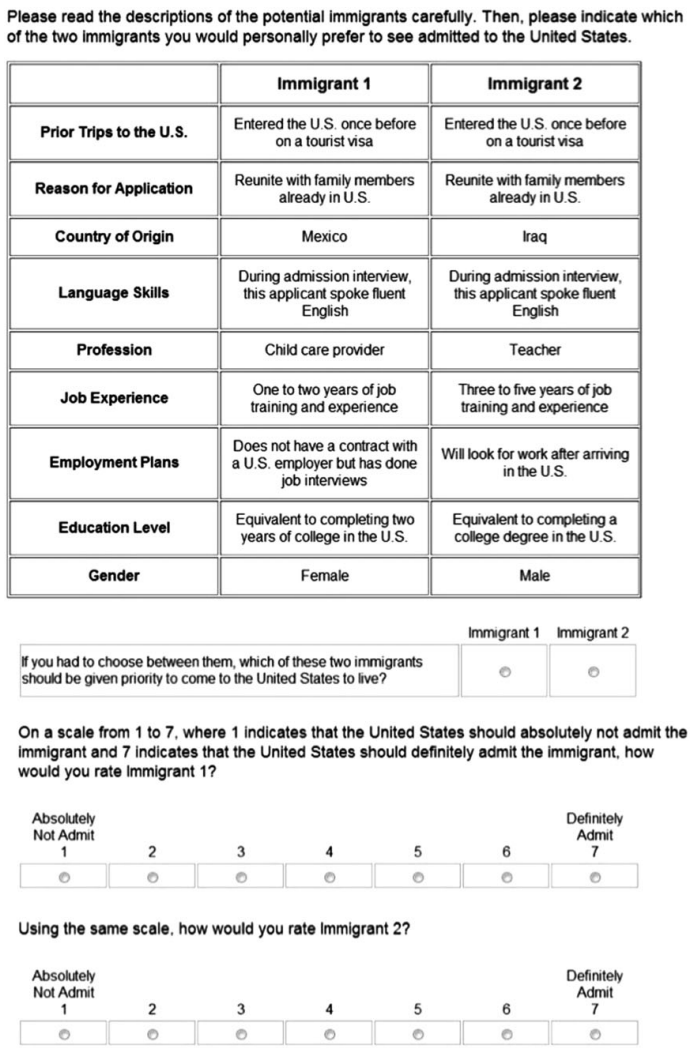
- Hainmueller et al 2013 - How do attitudes to immigrants depend on immigrant characteristics?
- Vary education, profession, language, gender, national origin, etc.
- Profiles
  - Attributes
    - Values
- Randomize attribute order to prevent bias

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 attribute order to prevent bias
- Treatment is the **combination** of attributes the respondent sees

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 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

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*.

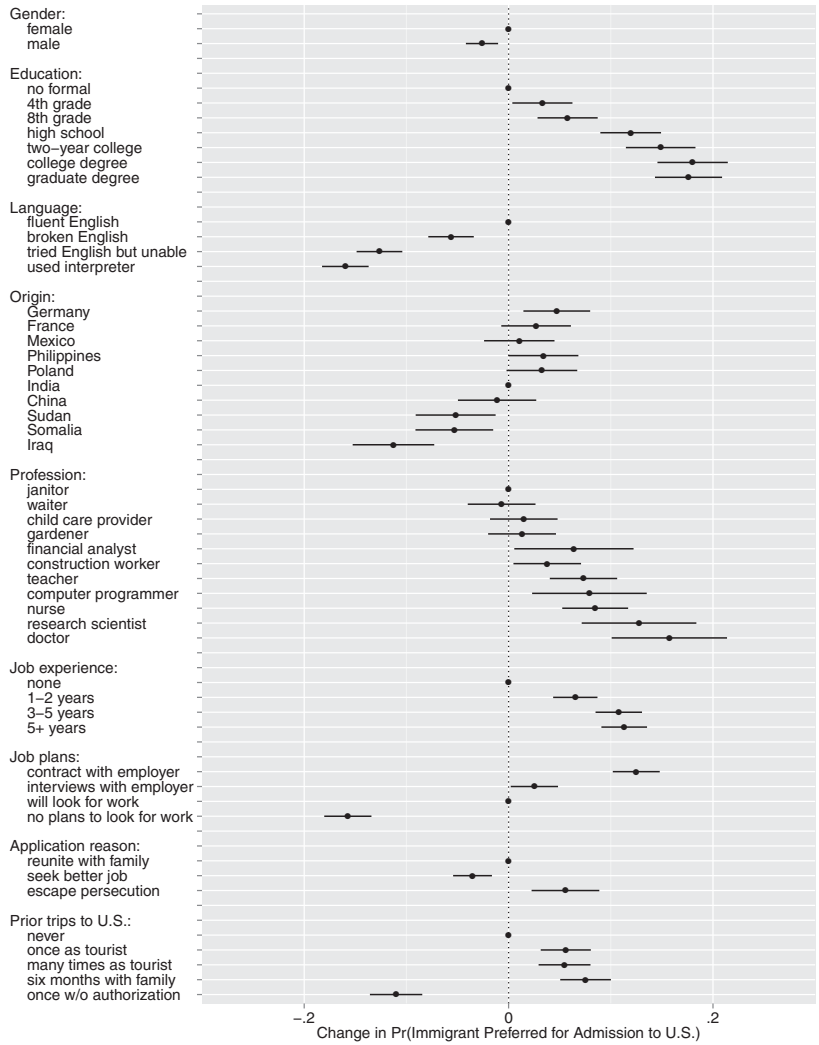


Fig. 3 Effects of immigrant attributes on preference for admission. This plot shows estimates of the effects of the randomly assigned immigrant attributes on the probability of being preferred for admission to the United States. Estimates are based on the regression estimators with clustered standard errors; bars represent 95% confidence intervals. The points without horizontal bars denote the attribute value that is the reference category for each attribute.

Conjoint Survey Experiments

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

Conjoint Survey Experiments

- Estimating results uses a simple regression of respondent choices on profile attribute-values
- But each specific profile (treatment) arises too rarely to make comparisons of individual attribute-values

Conjoint Survey Experiments

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

Conjoint Survey Experiments

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

Conjoint Survey Experiments

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

Conjoint Survey Experiments

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

Conjoint Survey Experiments

- We're still assuming people try to answer honestly

Conjoint Survey Experiments

- We're still assuming people try to answer honestly
- The ordering of attributes and profiles does not matter

Conjoint Survey Experiments

- How realistic are the responses?

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a behavioural measure; nothing at stake

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a behavioural measure; nothing at stake
  - Still some social desirability bias?

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a behavioural measure; nothing at stake
  - Still some social desirability bias?
  - Not like real-world preference-formation process
    - Stated preferences vs. Revealed preferences

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a behavioural measure; nothing at stake
  - Still some social desirability bias?
  - Not like real-world preference-formation process
    - Stated preferences vs. Revealed preferences
- Hainmueller et al 2014 - compare conjoint responses to a Swiss referendum

Conjoint Survey Experiments

- How realistic are the responses?
  - Not a behavioural measure; nothing at stake
  - Still some social desirability bias?
  - Not like real-world preference-formation process
    - 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)

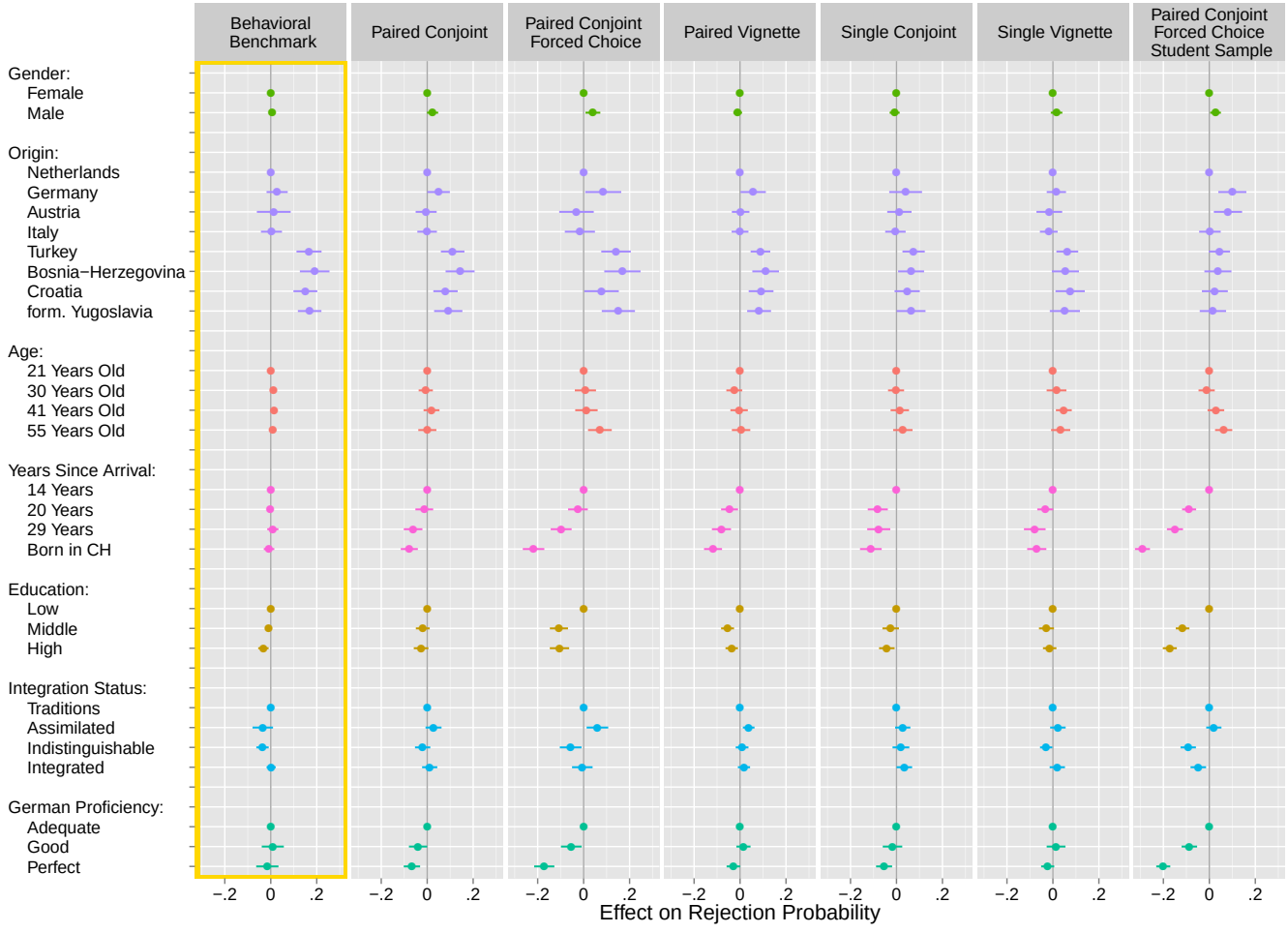


Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Conjoint Survey Experiments

- But note the conjoint method still hugely under-estimated the overall rejection rate
- 21% versus 37% in reality

Section 4

Generalizability

Generalizability

1. Generalizability of our Sample
2. Generalizability of our Context
3. Generalizability of our Treatment