

# Survey Experiments in Social Science

## Session 3 – Experiments in surveys

Federico Vegetti  
University of Turin

---

University of Newcastle

# Surveys

A survey is a method of investigation that uses **interviews** to collect information about people's opinions and behaviors

Interviews can be *structured*, *semi-structured* or *unstructured*

Here we will be talking about **structured interviews**, also known as *standardized interviews*

# Survey Experiments

What are the characteristics of survey experiments?

1. A deliberate manipulation of
  - The **order** of the questions that are presented to the respondents, or
  - The **wording** of the questions that are presented to the respondents
  - Other types of stimuli (e.g. visual, acoustic) that are presented to the respondents
2. Random assignment of the respondents to the different conditions

# Advantages of Survey Experiments

Better external validity than lab experiments

- Although still low ecological validity

Quite easy to implement

- Professional software like Qualtrics, Limesurvey and many more have implemented functions to randomize the order or the appearance of questions

Considerably cheaper than lab or field experiments

- You pay for the sample
- It is easier to find the chance put an experiment in a collective survey than to run your own lab experiment
- Cheap options like Mturk or Prolific

# What to do with Survey Experiments

According to Sniderman (2011), three types of design:

## **Manipulative**

- Get people to do what they are not predisposed to do

## **Permissive**

- Allow people to do what they are predisposed to do

## **Facilitative**

- Provide people reasons to do what they are predisposed to do
- (tbh not really clear to me how this is different from other design types)

This classification is based more on the goals of the experiment, rather than the type of design

# Manipulative Design

Experiments designed to demonstrate how easily people can be *nudged* into doing something

**Nudge:** use indirect suggestions to prompt people to behave in a certain way

The goal of nudging is to achieve a given behavioral change without forcing people into it

- For instance, we want citizens to send their tax return forms
- We can send different letters to different individuals, including empty of pre-compiled tax returns, and with different versions of the accompanying letter
- We can compare which version of the letter prompted individuals to send their tax return form

# Framing Experiments

An example of the so-called manipulative design are framing experiments

**Framing** refers to how a certain message is presented to the respondent – not the actual content of the message

For instance, when presenting a policy to the citizens, if a *positive* frame is used (i.e. emphasizing the advantages or gains of the policy) citizens are more likely to accept it than when a *negative* frame is used (i.e. emphasizing the losses or disadvantages)

Framing experiments make a very large part of experimental research in political communication

# Manipulative Design

This type of experiments have the potential disadvantage to reproduce unrealistic or unlikely scenarios

Sure, we may elicit a response, but how likely is it to happen in real life?

For instance, policy frames are rarely uniform

The informational environment to which citizens are exposed is likely to be much more complex, so the actual influence may be minimal



# Permissive Design

Experiments designed to remove constraints that respondents may have in other survey conditions

The idea is that some specific attitudes and behaviors are difficult to disclose (hence to observe), and the experimental design can come to our help

The difference from the manipulative design is that here respondents are not supposed to change their attitude or behavior

Rather, they are supposed to reveal their “true” attitude (or behavior) without constraints

# List Experiments

- Goal: get respondents to express potentially questionable opinions, or admit potentially questionable behaviors, without forcing them to do so
- Example: we want to know to what extent people are racially prejudiced
- We can provide them with a list of items, and ask them how many items in the list do not like
- The control group has 3 items: *cats*, *losing my wallet*, *slow internet*
- The treatment group has 4 items: *cats*, *losing my wallet*, *slow internet*, *a black family moving in my neighborhood*
- We compare the average response between the two groups: the difference should reflect the share of racially prejudiced people

# Endorsement Experiments

- In this case, we ask respondents to what degree they support a certain policy
- To half respondents, we just show a short description of the policy
- To half respondents, we also tell that the policy is endorsed by a given actor
- The difference between in average support for the policy tells us how much people support the actor
- Hard to tell if they are **permissive** (people reveal their true attitude about the actor) or **manipulative** (people change their opinion on the policy depending on the actor endorsing it)

# Conjoint experiments

In political science we often want to analyze people's preferences about multidimensional choice options

- E.g. voting, preferences for policies, candidates, etc.

Multidimensional: usually the choice options differ from one another on more than one characteristic

E.g. a candidate can endorse a certain set of policies, can have some relevant traits, can appear more or less competent.

What counts more? Conjoint experiments help us answering this type of questions, in a causal inference framework

# Conjoint experiments

Technique introduced in the 1970s in marketing research to study consumer preferences

Has become very popular in recent years thanks to the spread of computer-administered surveys

In practical terms, conjoint experiments are factorial experiments with a lot of factors and factor levels

It typically consist of tables or vignettes that are presented to the respondents, with a lot of characteristics varying randomly

# Example in table format

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
<b>Prior Trips to the U.S.</b>	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
<b>Reason for Application</b>	Reunite with family members already in U.S.	Reunite with family members already in U.S.
<b>Country of Origin</b>	Mexico	Iraq
<b>Language Skills</b>	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
<b>Profession</b>	Child care provider	Teacher
<b>Job Experience</b>	One to two years of job training and experience	Three to five years of job training and experience
<b>Employment Plans</b>	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.
<b>Education Level</b>	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
<b>Gender</b>	Female	Male

Source: Hainmueller et al., *Political Analysis*, 2014

# Types of outcome

The two most common types of conjoint experiments differ in the outcome that is measured

## 1. Choice outcomes

- Respondents are given two profiles, and are asked to choose between them
- More similar to real-world choice tasks

## 2. Rating outcomes

- Respondents are given one or more profiles, and are asked to rate it on a scale
- More direct measure of preferences

Researchers tend to measure both in the same survey

# Example in table format

	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?	<input type="radio"/>	<input type="radio"/>

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	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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

Absolutely Not Admit							Definitely Admit
1	2	3	4	5	6	7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Source: Hainmueller et al., *Political Analysis*, 2014



# Average Marginal Component Effect

The quantity that we estimate in conjoint experiments (the “average treatment effect”) is called Average Marginal Component Effect (AMCE)

In simple terms, the AMCE is the marginal effect of an attribute (e.g. gender = female) averaged over the joint distribution of the remaining attributes

It is similar to estimate the coefficient of an IV in a multiple regression, however in this case all the other attributes are randomly assigned, hence they can be assumed to be identical on average

# Many attributes

The strength of conjoint experiments is in the face that we vary *many* attributes between profiles

If we just varied e.g. one attribute, all the other considerations that might affect the outcome will remain out of our control

Recall:  $Y = T + X + U$

Where  $T$  is the treatment,  $X$  are observable and  $U$  are unobservable factors that influence  $Y$  beyond the treatment

In conjoint experiments, we minimise  $U$  and make sure that  $X$  is independent from  $T$

# How many attributes?

When deciding how many attributes to vary, we face a trade-off:

1. Many attributes are good, but too many might cause survey satisficing
  - People are cognitively overloaded, hence they might put low or no effort in the survey
2. Too few attributes might lead to masking
  - Respondents perceive a correlation with an observed and an unobserved attribute
  - E.g. you “manipulate” education, but people actually think of “age”

# How many tasks?

Another concern is whether respondents are given too many evaluation tasks (e.g. how many choices they are given in the survey)

Too many tasks might be also cognitively demanding, leading the respondents to provide low quality responses

But what is too many?

# Number of attributes and tasks

Research by Bansak and colleagues suggests that:

1. We can use a dozen or more attributes
  - Very low risk of satisficing
2. We can ask the respondents to perform 10 or even more tasks
  - No evidence of fatigue on the side of the respondent

# Sample size

One common concern with survey experiments is that many attributes that vary on many levels require a very large number of observations

However, in conjoint experiments we are not comparing specific profiles, but rather estimating the effect of every single component holding all the others constant

This implies that we do not have to have enough observations for every combination of factors

Still, the more observations, the more likely that the effect will be visible

# Effect heterogeneity

What if we think that the effect of some attributes can vary between different groups of respondents?

E.g. a candidate's competence might be perceived as more or less important by respondents of different levels of education

To do so, we need to collect some pre-treatment characteristics of the respondents, and then:

1. Estimate the AMCE separately for different groups
2. Interact pre-treatment and treatment variables (less common)

# Interactions

Usually, the attributes in conjoint experiments are independent from one another

However, this might create highly implausible or even impossible scenarios (e.g. an 18-yo person with a university degree)

To take care of this, it is possible to make the appearance of the levels of some factors conditional on other factors

In such a case, we will have to introduce in the regression model to estimate the AMCE an interaction between the 2 variables



# Issues with Survey Experiments

While easy and cheap to implement, survey experiments present a number of issues

- The duration of the effect that they elicit is likely to be limited, and often it is never observed
- Respondents are likely to get to the survey being already pre-treated
- Given the presence of multiple experiments and items in the same surveys, there may be cross-effects

# Survey Experiments and Effect Duration

Since many survey experiments are about attitude change, there is always the problem that the attitude change may disappear over time

In some cases, researcher conduct follow-up interviews days, weeks or months after the experiment

The effect often vanishes

In most of the cases, researchers do not even check for the duration of the effect

# Survey Experiments and Effect Duration

One could argue that the effect elicited in the survey is just an *instance* of the effect in the real world

The experiment may not have the pretention to perfectly mimic reality, but rather to discover *what* changes people's attitudes and *how*

More practically, it is often just not possible to have a follow up interview with the same subjects

# Survey Experiments and Pre-Treatment

The paradox of survey experiments and pre-treatment:

If the effect that we want to investigate is relevant, then it is likely to have occurred in the real world

Hence, respondents are likely to have been exposed to the treatment already

If the effect never occurred, there would be no motivation to do research

If the effect has occurred but it has vanished already, then its relevance is questioned as well

# Survey Experiments and Pre-Treatment

Pre-treatment is a kind of noncompliance

Hence, in most of the cases, it is likely to *deflate* our treatment effect, rather than inflating it

However, it is important to discuss the possibility that respondents have been pretreated when presenting experimental findings, and for how many of them this is likely to have happened

# Survey Experiments and Cross-Effects

A related problem, although more specific of the survey method, is the one of **cross-effects**

Previous manipulations or items in the survey may have a systematic effect on our experiment

Given the way survey experiments work in practice (i.e. a lot of people contributing to one survey with their own experiments and questions) you will encounter this

The best way to deal with this is to randomize the order of the manipulations and the items

# Is a Control Group Necessary?

One of the criticisms made to framing experiments is that, when comparing different frames, they lack a *baseline category* or *control group*

The control group in this case would be the information presented without any frame

In other cases, it is not obvious which group would be the control group

(think of the conjoint experiments)

# Is a Control Group Necessary?

In general, the control group shows how the phenomenon under investigation works in nature

- Without this information, it is very difficult to tell what is the effect of the experiment

In many experiments, the control group can also help our case with external validity

- If we show that the sample responds in a plausible way to a condition without additional stimuli, that is one indicator that our data are of good quality

It really depends on your theoretical research question

- If you only take advantage of some aspects of experiments, like randomization, then what you contrast with what may be different than in the classic experimental design