A primer on Survey and Experimental Methods for Public Economics

Stefanie Stantcheva

National Tax Association 118th Annual Conference on Taxation

SOCIAL
ECONOMICS
LAB

Social Economics Surveys and Experiments as a Key Research Tool

Large scale surveys that go in depth into people's minds and "listen to them."

Surveys have been used for a long time for statistics. Some variables are now better measured in high-quality administrative data (like income, family situation, employment, etc.)

Yet, some things are invisible in data other than survey data (even great data!): perceptions, attitudes, knowledge, and views.

In principle, one could specify complete structural model of beliefs and other "invisible" factors, but requires many assumptions identifying variations absent in data.

You are *creating the process that generates the data*. You can create your own controlled and identifying variation.

For the results to be reliable, it is critical that these surveys are well-designed, carefully calibrated, and deployed on appropriate samples.

Outline

Outline

There are many ways to build your sample

First question: what kind of samples do you need for your research question?

A useful notion: “sampling for range," the idea that your sample should be diverse in terms of conceptually important variables.

Channels you could use to build your sample:

- Nationally representative panels (e.g., *Understanding America Study*)
- Commercial survey companies which use quota sampled panels (e.g., *Qualtrics*)
- Commercial survey marketplaces (e.g., *Lucid*)
- Convenience samples (e.g., undergrads at your institution)
- Online work platforms (e.g., MTurk)
- Targeted groups from specific pools (e.g., employees at a firm)
- Government or institutions’ surveys (e.g., Survey of Consumer Expectations)

How online survey respondents compare to target populations

Target population: population of interest

Sampling frame: all the people in the population you can potentially sample

- Coverage error: difference between sampling frame and the target population. For example, in online surveys, cannot survey people who are not online

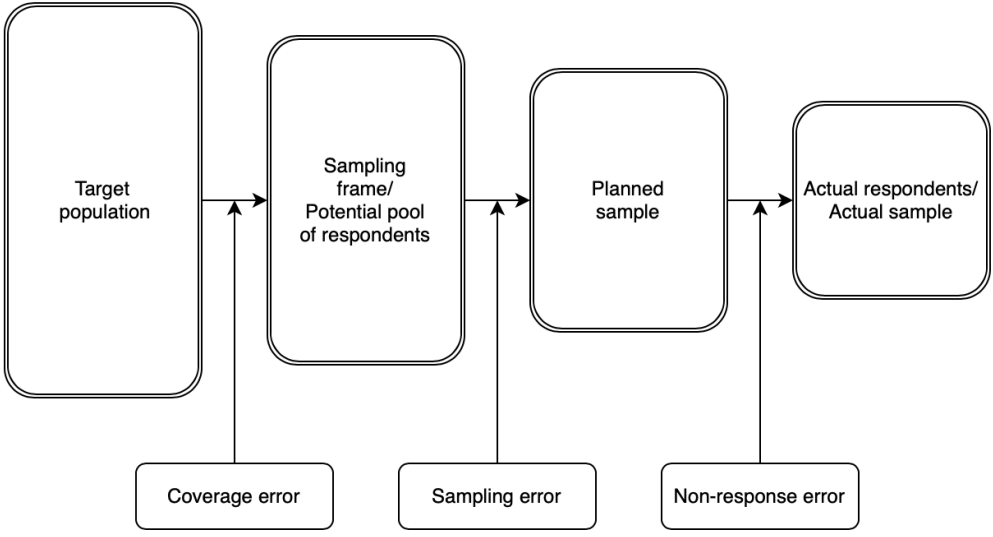
Planned sample: all people you want to complete your survey

- ***Sampling error***: difference between planned sample and the sampling frame

Actual sample: the people who end up taking your survey

- ***Non-response error***: difference between the target sample and the actual sample. E.g., due to not seeing invite
  - *unit non-response bias*: difference between respondents who start survey and those in planned sample
  - *item non-response*: when respondents start survey but some answers are missing. Includes *attrition* – respondents dropping out before finishing.

Survey errors and selection into online surveys



Selection: online surveys have advantages compared to in-person, phone, or mail surveys

1) People can take surveys at their convenience

2) Mobile technologies are convenient, so may encourage some people who would not otherwise want to fill out questionnaires or answer questions on the phone to take surveys

3) Can reach people otherwise difficult to reach (e.g., people in remote areas, people who move a lot)

4) Offer a variety of rewards, so can appeal to a broader group of people

National representativeness of online surveys varies across characteristics and country types

In the US and other high-income countries,

- Online samples can offer good representation of a broad income spectrum ($25,000 - $100,000), but very poor or rich
- Respondents skew more educated, more white, and somewhat more Democratic
- Respondents from larger urban areas and urban clusters are overrepresented

In middle-income countries, could be considered online representative rather than nationally representative.

Important to critically assess sample in light of survey methods and topic before generalizing

In particular, non-probability sampling (e.g., the quota sampling by survey company) has risks in terms of representativeness

Advice for recruiting respondents: surveys should be short, have good design, and appear legitimate

Reduce the perceived costs of taking the survey. Specify (ideally short) length of survey.

Use simple language and a friendly visual design. Make sure everything is easily readable (on mobile, too).

Do not reveal too much about the identity of the surveyor. Tradeoff between revealing more about yourself and providing info to make respondents feel confident. Revealing more can lead to bias, but can also provide legitimacy. Some info may be required by IRB.

- Ex: difference between “We are a group of non-partisan academic researchers" vs. “We are a group of faculty members from the Economics Department at Harvard and Princeton."

Appear legitimate and trustworthy. Provide contact information, consider IRB requirements, take into account rules such as the GDPR in the EU, and reassure respondents about complete anonymity and confidentiality.

Advice for recruiting respondents: Avoid selection bias, specify benefits, and motivate good response quality

Provide limited information about the purpose of the study. Some information is needed, but avoid revealing too much about actual research topic to avoid selection.

- "This is a survey for academic research" may be enough
- "This is a survey for academic research on immigration" will likely lead to some selection
- Never reveal purpose or intent of study (e.g., "We are interested in how people misperceive immigrants")

Specify some possible benefits of the survey. These could be for research and society broadly or for the respondent themselves

Warn against poor response quality. If appropriate, inform respondents that careless answers may be flagged and their pay may be withheld.

Attrition

Attrition rates tend to range between 15% to a bit above 30%, depending on platform and survey length

It is good practice to report detailed statistics on attrition:

- your total attrition rate with a clear definition
- your attrition rate at key stages in the survey
- "differential attrition:" correlations of attrition with respondent characteristics ($\Rightarrow$ avoid revealing topic too early and have people drop out before you know who they are)

How to prevent attrition:

- smooth respondent experience
- a shorter survey and good incentives
- avoid irritating respondents by having too many attention checks, personal questions, or complex questions

Correcting for non-response bias (selection and attrition)

Data can be missing for different entries in the survey:

- specific entries (item non-response)
- all entries (unit non-response)
- all entries after a given point (attrition)

Data can be missing to different degrees of randomness (at random, conditional on observables, not random).

Data is very rarely missing truly at random, but does not mean selection is always serious problem. Appropriate corrections depend on your goal and statements you are trying to make.

Four adjustment methods:

1) re-weighting observations
2) explicitly model selection or attrition into the survey
3) bound the effects of interest (rather than provide point estimates)
4) impute missing data directly (most often used for item non-response)

Addressing non-response bias: weighting methods

Cell-weighting. Sorting respondents and non-respondents into cells and reweight so they match the population

Related weighting methods. Include raking and generalized regression estimation

Logistics regression weighting and inverse probability weighting methods predict the probability of responding or completing the survey based on auxiliary information.

Standard errors. You must account for weighting when computing your standard errors

Addressing non-response bias: model-based approaches

Model-based approaches tackle attrition and sample selection parametrically. They explicitly model the selection or attrition process and do not need to assume that they are random, conditional on observables.

Models like in Heckman (1979) require finding an instrument that affects selection or attrition, but not the outcomes of interest.

- Could be randomized variation in survey process, e.g., number of times respondent was contacted or rewards offered
- If you have control over these parameters, can think ahead to their use later in your analysis

Addressing non-response bias: bounding methods

The worst-case approach by Horowitz and Manski (2000) imputes information using min and max possible values of outcome variables and bounds population parameters with almost no assumptions.

> Bounds can be wide and non-informative, but are useful benchmarks (esp. for binary variables)

Lee bounds. Lee (2009) proposes a method to bound treatment effects by estimates when there is differential attrition between treatment and control groups. The bounds are estimated by trimming a share of the sample either from top or bottom

> To apply these bounds, treatment has to be randomly assignment, and treatment assignment should only affect attrition in one direction

Addressing non-response bias: imputation methods

Imputation methods are non-parametric techniques to fill in missing data

Hot deck imputations replace missing values with a random draw from some "donor pool," which are values of the corresponding variable for responders that are similar in some way

Regression-based imputations replace missing values with predicted values from a regression of missing variable on observed variables for respondent. Little and Rubin (2002) suggest that imputations need to be:

- conditional on observed variable
- multivariate
- randomly drawn from predictive distributions

Outline

Ex ante methods to check for attention

First, need to collect “meta-data" (e.g., time spent on survey screen) to diagnose issues with attention and carelessness.

“Screeners" are questions specifically designed to detect inattentive answers and can be structured in different ways:

Logical questions require logical reasoning and have a clear, correct answer (e.g., *“Is the sky blue or green?”*)

Instructional manipulation checks look like regular survey questions but instruct the respondent to give a certain answer

Factual manipulation checks are questions asked after experimental treatments that have correct answers

Screeners can also annoy respondents and increase attrition rates, so should be used sparingly and strategically

Ex post data quality checks

After identifying potentially problematic respondents, could check the robustness of your results to including vs. dropping them. You Can also flags for different degrees of carelessness

Consistency indices are measures that match items supposed to be highly correlated. You can then check whether they are correlated (for example, the questions “Are you cold?" and “Are you hot?")

Response pattern indices detect patterns in consecutive questions

Outlier indices attempt to spot outlier answers

Honesty checks or self-reported attention are direct questions that ask the respondent about their interest or attention

Time spent on the survey can be used to determine whether someone who spent too little or too much time on a survey should be included.

How to reduce and test for survey fatigue

The erosion of respondents' focus and attention over the course of the survey is an important concern

Reducing survey fatigue

- Good design is critical.
- Length of survey is important.
  - No hard rules
  - Interesting, well-designed surveys may foster more engagement than shorter, poorly-designed, or boring ones

Testing for survey fatigue

- You can test for patterns of carelessness (though not always conclusive)
- You can randomize order of survey blocks to see if respondents who (randomly) saw a block later spend less time on it or show more carelessness

Outline

General advice (1/5)

Types of questions:

- **Closed-ended questions**: have a fixed set of answer options.
  - Nominal (without natural ordering, e.g., *"What is your marital status"*)
  - Ordinal (with some ordering, e.g., *"Do you support or oppose this policy?"*)
- **Open-ended questions**: have an open answer field of varying lengths (e.g., *"What comes to mind when you think about income taxes?" [empty text field]*)
- **Hybrid questions**: are close-ended questions with an open-ended answer choice (e.g., *"Other (please specify):" [empty text field]*)

Each question needs to ask about one thing at a time, with everything else being held as equal as possible

General advice (2/5)

Write precise questions. To avoid misinterpretations (which leads to measurement error), avoid the following kinds of questions:

- Double-barreled questions, i.e., questions that ask about two things simultaneously.
  - Ex: *"Do you support or oppose increasing the estate tax and the personal income tax in the top bracket?"*
- Vague questions.
  - Ex: *"Do you support or oppose raising taxes on the rich?"* It is probably what you mean by "rich" and which taxes you have in mind.

Allow for a respondent to answer that they do not know or are indifferent

General advice (3/5)

Use simple language.

- Know your audience (e.g., difference between adults and teenagers)
- Do not use jargon or undefined acronyms
- Do not use double negatives or negatives that are difficult to understand
- Eliminate all unnecessary words and keep questions short

Adapt your question flow to your respondents. Make sure questions apply to all respondents. If they do not, either

- create survey branches based on respondent characteristics
- ask contingent questions (e.g., do not ask an unemployed person about their job. Rather, can ask *"If you currently do not have a job, please tell us about your last job."*)

Example: adapting the question flow to your respondents

We will now ask you what you think the before tax income for **P50** was in 2017 for the groups below which you are a part of. The first slider shows your answer from the previous question. You can use the sliders to select what you think the income was for P50 for the different groups of **people who were born the same year as you**.

P50 for people **born in 1970**

33.000

P50 for **women** born in 1970

20.000

P50 for people who also lived in **Bornholms eller Christiansø municipality**

20.000

P50 for people who also had the educational level **Kort videregående uddannelse**

20.000

General advice (4/5)

Avoid leading questions that nudge the respondent in one direction.

- Avoid agree/disagree and true/false questions (acquiescence/agreement bias)
- Avoid judgmental or emotionally charged words
- Do not ask sensitive or private questions unless you have to
- Avoid giving reasons for a given questions

  Ex: "*Do you support or oppose higher taxes so that children can have a better start in life?*"
- Consider including a counter-biasing statement to signal neutrality.

  Ex: "*Some people support very low levels of government intervention in the economy while others support very high levels. How much government involvement do you support?*"
- When asking either/or questions, state both positive and negative sides

  Ex: "*Do you favor or oppose increasing the tax rate in the top bracket?*"

Do not force responses except when answers are needed to screeners at survey start. However, you can "prompt" for responses e.g., with a pop-up

General advice (5/5)

Use simple question formats unless question requires more complexity

- Checkbox questions: respondents select one or more options
- Radio button questions: respondents select one option
- Slide questions: respondents select an answer by moving a slider. The benefit is that it yields a more fine-grained and continuous answer
- Ranking questions: respondents rank options. Can be cumbersome and should only be used for actual rankings

Provide informative error messages. For example, “Your answer is invalid" is not helpful; “Please only enter integer numbers" is more useful

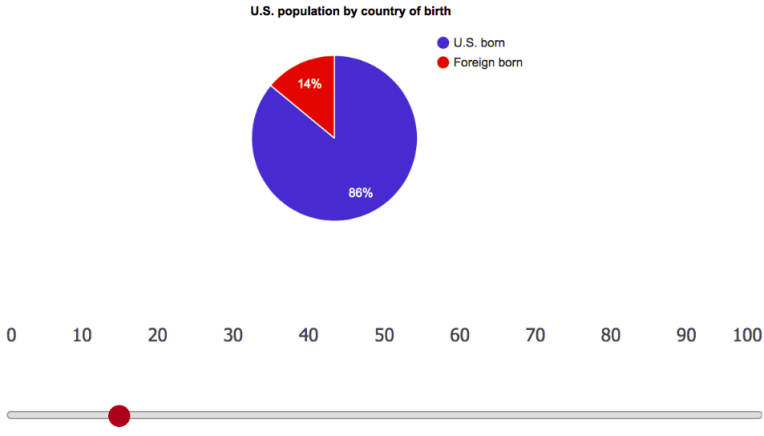
Look for precedent. Review the literature and existing survey questions.

Pre-test. Test the questionnaire on a wider, non-expert audience and ask for feedback.

Include feedback questions. Feedback entry fields at the end can be targeted or general and provide other useful insights

Example slider question: share of immigrants from Alesina et al. (2022)

The pie chart below represents all the people currently living in the U.S. Out of all these people currently living in the U.S., how many do you think are legal immigrants? Move the slider to indicate how many out of every 100 people you think are legal immigrants.



Benefits of open-ended questions

Allow researchers to elicit people's views/concerns without priming them with a given set of answer options

Helpful in exploratory work before writing a complete survey

Avoid leading and priming respondents

In context of information experiments, can be used to validate answers to other questions

Best practices for open-ended questions

Because more time-consuming than close-ended ones, **motivate respondents to answer** (e.g. "*This question is very important to understanding tax policy. Please take your time answering it.*")

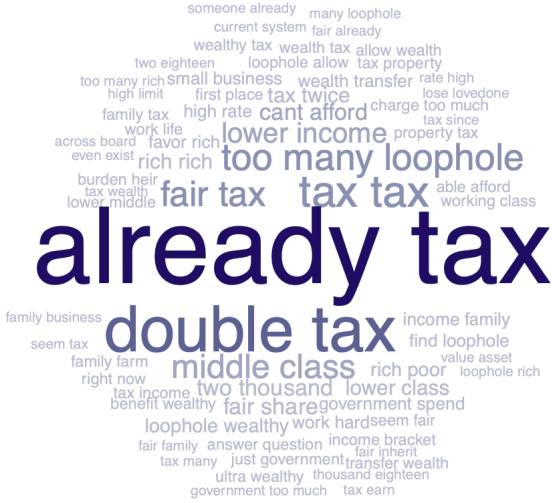
To encourage extended responses, could **add follow-up questions on next screen** (e.g., "*Are there any other reasons you can think of?*")

Use them sparingly, place near beginning (while respondents are less fatigued)

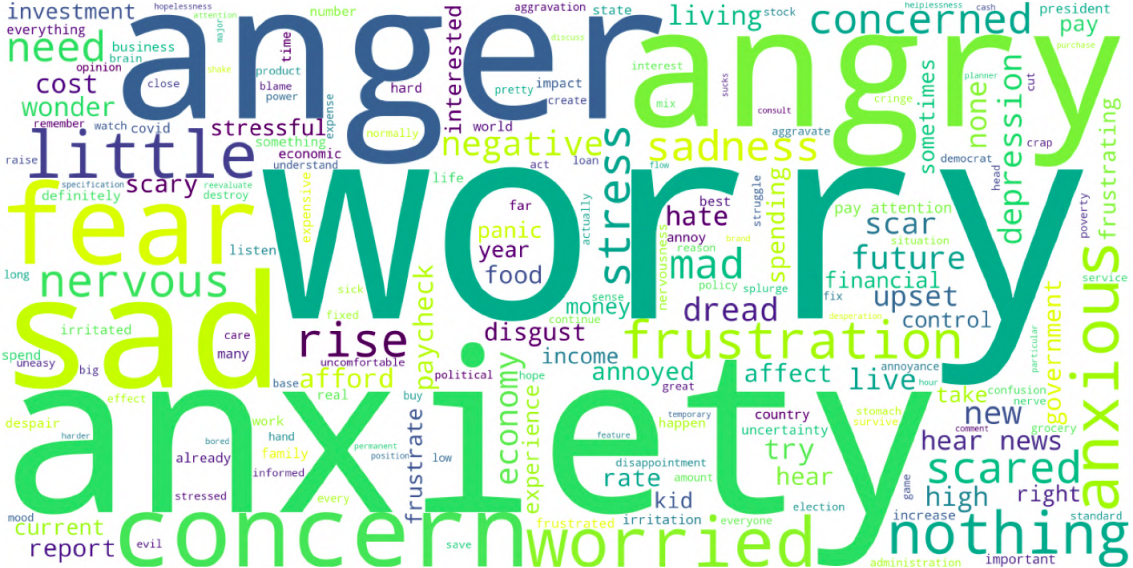
Specify what type of answers you are looking for (e.g., "*Please think of several reasons...*")

Adapt visual format to type of answer you need (e.g., if you are looking for a dollar amount, put a $ next to the box.)

Word cloud for "What is the biggest problem with the estate tax?"



Word cloud for "When hearing rising inflation I feel..."



Close-ended questions

There are two main types of ordinal closed-ended questions: qualitative and quantitative. They can have different answer options:

- more vague response options (e.g., never, rarely, sometimes, always)
- responses using natural metrics (e.g., once a week, twice a week, etc.)

Include multiple measurements of important variables in your survey, especially in cases where vague qualitative options may have different meanings across people

Use exhaustive answer options that span all possible reasonable answers.

Make sure answer categories are non-overlapping.

> For example, answer options to "*What should the marginal tax rate for incomes over $1 million be?*" should have a non-overlapping scale such as [0% to 19%, 20% to 29%, etc.] rather than an overlapping one like [0% to 20%, 20% to 30%, etc.]

Use a reasonably small number of answer options. Keep the answer list as short as possible, with exceptions e.g., when asking about ethnicity or gender, where you should aim to be inclusive

Specific advice for nominal closed-ended questions

Multiple choice questions: “forced-choice” often preferable over "check-all-that-apply"

- Forced-choice: ask item by item and require respondents to judge them independently. They generally lead to respondents thinking more carefully about the options and can address order effects in questions.
- Check-all-that-apply: list items simultaneously, ask respondents to select at least one item.

Check-all-that-apply question

Which of the following policies do you think could reduce inequality? Check all that apply.

- ☐ Job re-training programs
- ☐ Higher income taxes
- ☐ Higher minimum wage
- ☐ Free early childhood education
- ☐ Anti-trust policies
- ☐ Unemployment insurance

Forced-choice question

Do you think of the following policies could reduce inequality?

	Yes	No
Job re-training programs	☐	☐
Higher income taxes	☐	☐
Higher minimum wage	☐	☐
Free early childhood education	☐	☐
Anti-trust policies	☐	☐
Unemployment insurance	☐	☐

Specific advice for nominal closed-ended questions (cont.)

Choose direct and adapted scales.

Question with indirect scales
To what extent do you agree or disagree that the income distribution in the US is fair?
- ☐ Strongly agree
- ☐ Agree
- ☐ Neither agree nor disagree
- ☐ Disagree
- ☐ Strongly disagree

Construct-specific questions
How fair or unfair do you think the distribution of income is in the US?
- ☐ Very unfair
- ☐ Somewhat unfair
- ☐ Neither fair nor unfair
- ☐ Somewhat fair
- ☐ Very fair

Use natural metrics. For example, use “last week" or “last month" rather than “regularly" or “often" when asking about how many times a respondent did something

Label all options in an answer scale, not just extremes.

Remove numerical labels unless they have true meaning. For example, do not add numeric labels to answer options such as “1 = strongly agree," “2 = agree"... “5 = strongly disagree." This is distracting and potentially misleading

Measurement issues (1/2)

Quantitative and technical issues. Many studies will require new, creative, and sometimes complex questions. For example, respondents may be asked to select their position on a ladder representing the income distribution using a slider. For these questions:

- should pilot them them several times and check for understanding
- automate your survey code so that, e.g., percentages add up to 100.
- good visual representation is key

Point beliefs vs. probabilistic beliefs. Point beliefs are easier for respondents to understand, but do not allow respondents to express uncertainty. Probabilistic beliefs may be harder to understand and yield noisier results, but allow respondents to express uncertainty.

Example of a complex question with visual representation from Alesina, Stantcheva, and Teso (2018)

We would now like to ask you what you think about the life opportunities of children from very poor families.

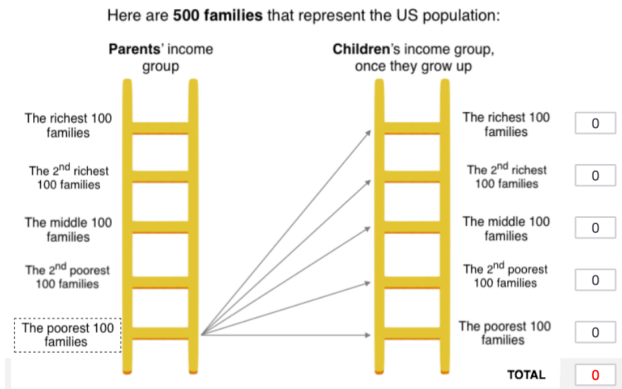
For the following questions, we focus on 500 families that represent the U.S. population. We divide them into five groups on the basis of their income, with each group containing 100 families. These groups are: the poorest 100 families, the second poorest 100 families, the middle 100 families, the second richest 100 families, and the richest 100 families.

In the following questions, we will ask you to evaluate the chances that children born in one of the poorest 100 families, once they grow up, will belong to any of these income groups.

Please fill out the entries to the right of the figure below to tell us, in your opinion, how many out of 100 children coming from the **poorest** 100 families will grow up to be in each income group.

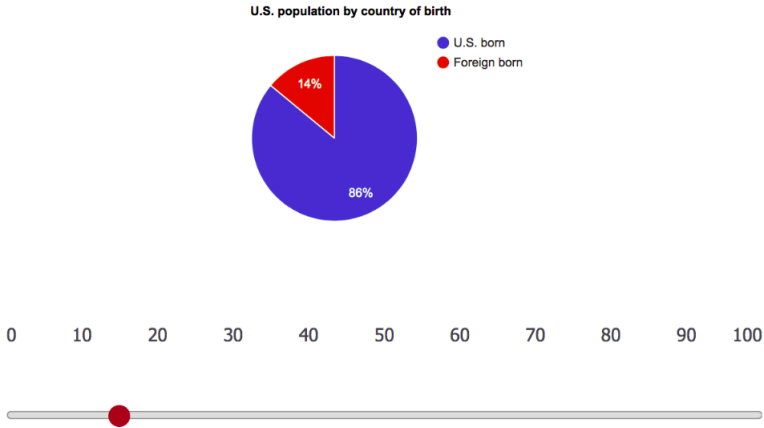
From our experience, this question will take you at the very least 1 minute to answer.

Please note that your entries need to add up to 100 or you will not be able to move on to the next page.

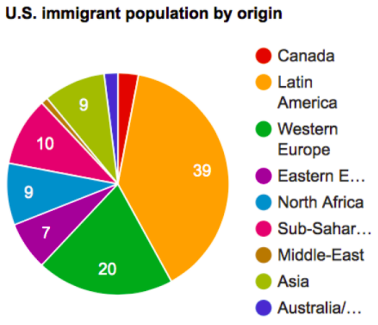


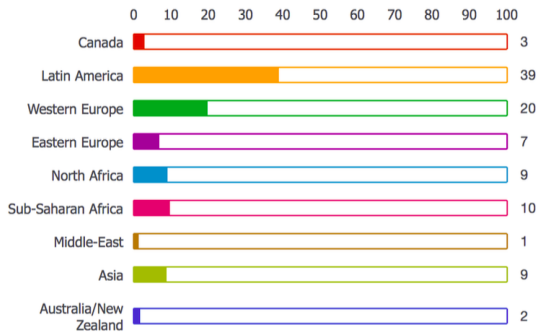
Eliciting perceptions on share of immigrants

The pie chart below represents all the people currently living in the U.S. Out of all these people currently living in the U.S., how many do you think are legal immigrants? Move the slider to indicate how many out of every 100 people you think are legal immigrants.



Eliciting perceptions on origin of immigrants





Eliciting beliefs about ideal income tax rates

The government currently raises a certain amount of revenue through the income tax in order to sustain the current level of public spending. In you view, what would be the fair split of the tax burden to sustain public spending?

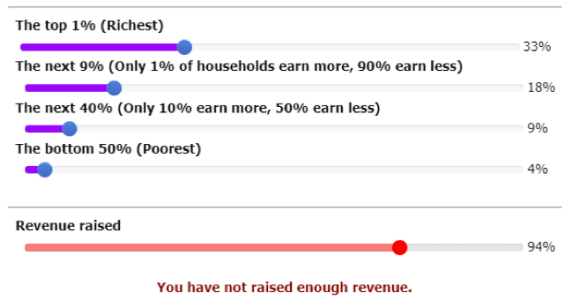
The income tax* rate is the percentage of your income that you pay in federal income tax. For example, if you earn $30,000 and you pay $3,000 in income taxes, your income tax rate is 10%.

Please use the sliders below to tell us how much you think each of the following groups should pay as a percentage of their total income.

While you adjust the four sliders for each group, the fifth bar at the bottom moves in order to show you how much of the current revenue you have been able to raise so far. The bar appears red as long as you have not raised enough revenue, or if you have raised more money than what is needed.

You will only be able to move to the next question when you meet the revenue target and the bar becomes green.

** We consider only the Federal income tax, which is a tax on household income. If you receive a regular paycheck, this tax is automatically taken out of your pay. When you file a federal tax return each year, you calculate the exact amount you owe, and you get a tax refund from the federal government if you paid more than you owe. To keep things simple, we do not include other taxes such as social security taxes, state income taxes or sales taxes.*



Measurement issues (2/2)

Multiple measurements. Perceptions and beliefs are hard to measure, so measurement error is a serious risk. One solution is to use multiple measurements of same variable.

The importance of benchmarking. Suppose you find respondents estimate a given variable quite incorrectly. Is this misperception specific to given variable, or is it the manifestation of a general bias?

- Ex: People starkly overestimate share of unemployed immigrants, but also overestimate unemployed non-immigrants. So when you are interested in one group, ask about another (in a random order).

Interpersonal comparisons and anchoring. Different interpretations may be common in cross-country studies or across respondents with different backgrounds. Anchoring vignettes or questions – short descriptions of hypothetical people or situations – may reduce problem.

Visual design (1/2)

Mobile version. Make sure all questions and overall layout display well on mobiles and a range of browsers

Fonts. Use darker/larger fonts for question stem and lighter/smaller ones for answer options.

Spacing. Use spacing to help create sets within a question

Standardize. Use a common graphic standard for all or most questions.

Emphasis and de-emphasis. Visual design can help guide respondent's attention.

Special instructions. Try to adapt your survey to avoid special instructions, but for questions that do require them, place them correctly

Optional or occasionally needed instructions. Reduce respondents' burden by separating elements that do not read from elements that are required.

Visual design (2/2)

Separate substantial and non-substantial answer options. Answers such as “I don’t know" should be at the end of the answer options

Simplify. Try to avoid visual clutter on the page. Do not place questions next to each other (like shown below)

Type of tax	To what extent do you think that an increase in this tax will hurt economic activity in the US?			Do you think this tax is fair or unfair?				
	Not Hurt at all	Hurt somewhat	Hurt a lot	Very fair	Fair	Neither fair nor unfair	Unfair	Very unfair
Personal Income tax	☐	☐	☐	☐	☐	☐	☐	☐
Property tax	☐	☐	☐	☐	☐	☐	☐	☐
Sales tax	☐	☐	☐	☐	☐	☐	☐	☐
Excise tax	☐	☐	☐	☐	☐	☐	☐	☐
Payroll tax	☐	☐	☐	☐	☐	☐	☐	☐

Question ordering

Ordering should be guided by three (sometimes conflicting) concerns:

- Respondents are often more engaged and less tired at beginning
- Questions that come at beginning may influence later responses
- Respondents form opinion about your survey at beginning and it's critical to capture their interest

Common advice is to ask exciting questions first

If you are worried about attrition, ask demographic info first

Sensitive questions should come later in the survey

Organize your survey logically and guide respondents

If you ask filter questions, (e.g. "*Are you self-employed?*"), ask all of these first and then ask the follow-ups (e.g. "*What is your income from self-employment?*")

Using monetary incentives and real stakes questions

Monetary incentives for truthful revelation can be used to reward accurate answers or particular behaviors.

Real stakes questions can be used to lend credibility to self-reported attitudes, behaviors, and beliefs

- Can include e.g., petitions or donations to charities related to the issue of interest

Making respondents "spectators" can help respondents internalize their self-reported choices. As spectators, they observe the actions or choices of other respondents (the "stakeholders") and then allocate rewards to them. Spectators' choices are then implemented with a certain probability.

Real stakes questions: multipe price list for a gun safe from Alsan, Schwartzstein and Stantcheva (2025)

Instructions

You will have to make several decisions by completing one list. Each row of the list will present two payment options:

The payment option **on the left** will involve you receiving **$5 as a bonus payment**, which you'll get if you select this option.

The payment option **on the right** will always involve getting a **discount on a RPNB Biometric Safe**.

The amount of the discount will increase from $5 to $90 as you proceed down the rows of the list. In the first row, you'd be able to purchase the RPNB Biometric Safe at a $5 discount (bringing the total price to roughly $100- $5 = $95) and in the last row you'd be able to purchase a RPNB Biomertric Safe at a $90 discount (bringing the total price to roughly $100- $90 = $10).

Your task is to decide which payment option you prefer on each row.

By participating in this survey, you are automatically entered into a lottery. If you win the lottery, one row from this list will be randomly selected. The payment option you select on that row would then be distributed to you free of charge.

Real stakes questions: multipe price list for a gun safe from Alsan, Schwartzstein and Stantcheva (2025)

Please indicate **which payment option you prefer on each row**.

(Note that you cannot click on the submit button until you have selected an answer.)

Which option do you prefer?

Bonus payment of $5 ○	A $5 discount, bringing the price of the RPNB Biometric Safe to $95 ○

Which option do you prefer?

Bonus payment of $5 ○	A $22 discount, bringing the price of the RPNB Biometric Safe to $78 ○

Which option do you prefer?

Bonus payment of $5 ○	A $39 discount, bringing the price of the RPNB Biometric Safe to $61 ○

Which option do you prefer?

Bonus payment of $5 ○	A $56 discount, bringing the price of the RPNB Biometric Safe to $44 ○

Which option do you prefer?

Bonus payment of $5 ○	A $73 discount, bringing the price of the RPNB Biometric Safe to $27 ○

Outline

Types and sources of bias

Types of bias

- moderacy bias
- extreme bias
- ordering bias
- acquiescence bias
- experimenter demand effect
- social desirability bias

Sources of bias

- the respondents' behavior (e.g., carelessness or social desirability bias)
- the content of the question (e.g., leading questions)
- the design of the questionnaire (e.g., the order of questions that can induce priming)
- the characteristics of the survey situation itself (e.g., experimenter demand effect)

Unintended question order effects

Why do question order effects occur?

- *"Cognitive-based order effects"*: include priming, carryover, and anchoring
- *"Normative-based order effects"*: include the respondent's wish to appear: evenhanded or fair, consistent, and moderate

Possible solutions

- No general solution, but be aware of effects when designing survey.
- Can randomize order of questions or blocks if it does not negatively impact survey flow.

Social desirability bias

Social desirability bias (SDB): comes from respondents' desire to avoid embarrassment and project a favorable image to others, which results in them not revealing their actual attitudes.

Prevalence depends on the topic, questions, respondent, survey mode, and social context.

Online surveys likely minimize SDB relative to other survey modes because there is no surveyor in front of respondent

- Revealed identity of surveyor, level of anonymity provided, and knowledge of what questions will be used for are the "social context" in online surveys

To address, can repeat reminders of anonymity.

Methods to address SDB (1/2)

Randomized response technique: uses a randomization device such as a coin flip for which only the respondent sees the outcome. Outcome determines which question respondent needs to answer or how they should answer. By knowing the probability of each random event, researcher can infer true proportion of socially undesirable behavior.

- Mirrored questions
- Forced-response questions
- Disguised response technique
- Unrelated question design

Cross-wise technique: A sensitive question is paired with a non-sensitive question, the answers to which have a known distribution. Respondents answer the two questions jointly.

- Cross-wise technique

Methods to address SDB (1/2)

The unmatched count technique or list experiment: there are at least two groups and each group is given a list of items. The control receives control items only while the treatment group receives the control list plus a sensitive item. Then, can compare difference in mean number of items between treatment and control. Can also be extended into a double list experiment

▸ List experiment ▸ Double list experiment

Face-saving language: frame the question in a way that normalizes the sensitive behavior or that includes face-saving alternatives to signal surveyor's acceptance of the behavior.

▸ Face-saving language

Acquiescence bias

Acquiescence: the tendency to answer items in a positive way regardless of their content, for instance, systematically selecting categories such as “agree,” “true,” or “yes"

Possible solutions:

- Do not ask ambiguous, unclear, or complicated questions.
- Avoid “agree/disagree," “true/false," and “yes/no" questions. Instead, ask questions with direct, item-specific scales. Otherwise, you can offer answer options that include all possible views.
- With bipolar scales, always use a balance scale with an equal number of positive and negative options.
- You can consider having two versions of fundamental questions in your survey, one with a negative statement (use sparingly).
- At a group level (e.g., within a country or age bracket), you can randomize who sees a given question and who sees its inverse.
- You can adapt the reduced-form and model based approaches discussed earlier

Experimenter demand effect

Experimenter demand effect (EDE): Respondents may think the experimenter wants a certain kind of response. Treatment status can affect this.

Possible solutions:

- *Anonymity*. Online surveys can relieve social pressure.
- Monetary incentives and real stakes questions
- *Obfuscated follow-ups* are follow-up studies with the same respondents and dependent variables, but without the respondent knowing the original and follow-up are related.
- *Obfuscated information treatments* try to obscure the purpose of the experiment, e.g., by giving respondents with extra info irrelevant to goal, asking questions about unrelated issues, and giving people an unrelated reason why they receive information.
- Design and question wording
- Hiding the purpose of experiment or study
- Measuring beliefs about the study purpose
- *Demand treatments* explicitly introduce (possibly randomized) questions that use explicit signals of the surveyor's wishes and use them to measure extent of EDE.

Outline

General considerations for survey experiments

Challenges

- Confounding by unobservable factors and pre-treatment contamination
- *Information Equivalence*: different respondents can interpret experimental intervention differently
- Whether information mimics real-world treatments or whether it is more abstract
- External validity based on sample composition and setting

Different types of survey experiments:

- **Information treatments**: correct or expand respondents' information by learning and updating.
- **Priming treatments**: activate certain concepts or mindsets and make some features more salient
- **Vignette designs and factorial experiments**: modify various attributes of choice behavior in a controlled manner to study impacts on judgment and behavior

General design choices

Need to make some decisions on design, especially:

- Will you use a between- or within-respondent design?
  - Between-subject: each respondent is only subject to one experimental condition
  - Within-subject: each respondent is subject to multiple experimental conditions. Difference is order in which conditions are administered.
- When will you measure your dependent variable: before, after, or both before and after treatment?
- In information experiments (and sometimes others), different variables to consider:
  - “first stage beliefs" (the belief or information that your treatment is trying to shift)
  - “second stage" (dependent variables influenced by first-stage ones)

Different types of experimental designs (Clifford et al., 2021)

Experimental design type	**Group**		**Time T_1**		**Time T_2**
Posttest	1			Treatment	O_1
	2				O_2
Prepost	1		O_1	Treatment	O_2
	2		O_3		O_4
Quasi-prepost	1		Q_1	Treatment	O_1
	2		Q_2		O_2
Within	1	Treatment	O_1		O_2
	2		O_3	Treatment	O_4

Notes: O = observation of the dependent variables (first and second stage)
Q = observation of variables closely related to the dependent variables.
T_1 = point in the survey where first measurement possibly taken.
T_2 = point in the survey where second measurement possibly taken.

Priming treatments

Priming treatments activate mental concepts/mindsets through subtle cues. Can be used to measure attitudes without respondent knowing what is being measured.

Mechanism and theoretical aspects. Priming can change salience of a primed concept a given moment. Experiments can focus on:

- Conceptual priming: activates mental concepts in one context to induce mental representations that are used in subsequent contexts
- Mindset priming: primes a given way of or procedure for thinking by having participant use that procedure.

Two channels through which priming works:

- accessibility: making some more features more salient and thus knowledge can be reactivated in memory task
- applicability: degree to which presented stimulus or stored knowledge is perceived as applicable to another context

Types of priming

Slanted questions: eliciting a certain response from respondents by asking them about a certain event or experience.

- For example, asking about the Wall Street bailout to prime respondents to think negatively about the government.

Order randomization/changing order of question

- For example, Alesina, Miano, and Stantcheva (2018) test effect of immigration perceptions on view on redistribution randomizing the order in which respondents are asked about each

Use of words/names with different connotations

- For example, asking about undocumented vs. unauthorized vs. illegal immigrants, as in Merolla et al. (2013)

Varying the illustrations and images shown alongside the information

- For example, priming people about the lifestyle of the wealthy by showing a picture of a mansion, as in Kuziemko et al. (2015), who provide information on the low share of people who are subject to the wealth tax.

Priming through images and videos

Information treatment with emotional priming from Kuziemko et. al (2015)

Besides the income tax, the government can also level the playing field with **the federal estate tax**.

The **Federal Estate Tax** (also known as the **Death Tax**) applies when a deceased person leaves **more than $5 million** in wealth to his or her heirs. Wealth left to a spouse or charitable organizations is exempt from estate tax.

Only 1 person out of 1000 is wealthy enough to face the estate tax.

Average Americans do not have anything close to $5 million in wealth, so the estate tax does not affect them and they can pass on their property to their children tax-free.

Eliminating the estate tax would allow the very richest families to pass down all of their wealth to their children tax-free. Hence, children of rich people would also start off very rich themselves.

Increasing the estate tax is a way to level the playing field between the children of wealthy parents and children of middle-class parents.

Information and pedagogical treatments: exogenously changing information sets of respondents

Allow for testing for effects of specific information on outcomes, studying impact of correcting misperceptions, and belief updating.

Quantitative information

- Precise, clear, and minimizes differences in interpretation across respondents, but may be harder to understand
- Ex: showing evolution of earnings gap between Black and white men, as in Alesina et al. (2021)

Qualitative information

- Sometimes more suited to a question
- Can make treatment more homogeneous when testing across different countries or settings
- Alesina et al. (2018) show respondents in five countries an animation on lack of social mobility in their country

Anecdotes, stories, and narratives

- Ex: showing respondents an animated video about differences in opportunities of a Black and white child, as in Alesina et al. (2021).

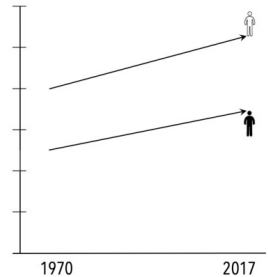
Example of quantitative information treatment: evolution of earnings gap between Black and white men from Alesina, Ferroni, and Stantcheva (2021)

(D) Black-white Earnings Gap in 1970

In 1970, a black person would on average earn only two thirds as much as a white person.
That is, for every dollar a white person earned, a black person would earn 63 cents.

(E) Black-white Earnings Gap Today

Over time, the earnings of white and black people have grown.
But the gap in earnings between black and white people has not been closed at all over the years.



Example of a quantitative information treatment: rank in the income distribution from Kuziemko, Norton, Saez and Stantcheva (2015)

Please enter your annual household income* in the box below:

$ 25000

39% of US households earn less than your household

We now encourage you to move the blue slider above (by clicking on the line) to explore the US income distribution on your own and to answer the questions below.

50% of households earn less than **$33,800** .

Example of a quantitative information treatment: income if growth had been equally spread from Kuziemko, Norton, Saez and Stantcheva (2015)

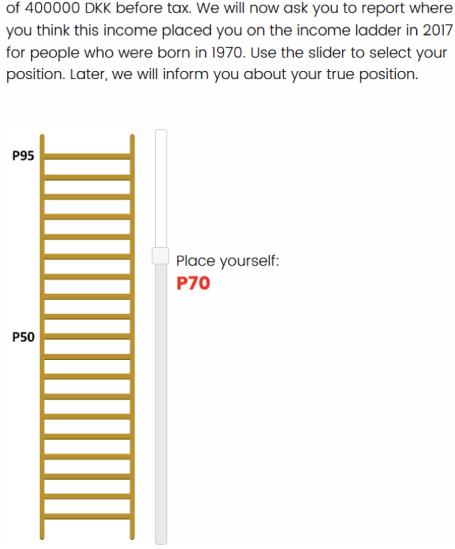
Income Inequality has increased dramatically in the United States since 1980.
Incomes of poorer and middle-income families have grown very little while top incomes have grown a lot.

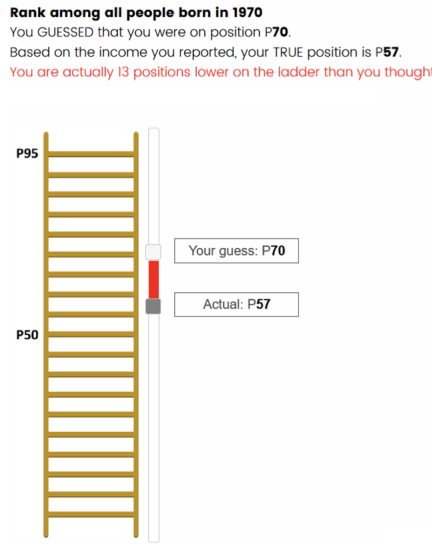
How would YOU be doing if inequality had not increased?

The slider below shows how much each group would make if incomes had grown by the same percentage since 1980 for all groups: the poor, the middle class, and the rich. Use the slider to answer the questions below.

A household making $30,400 today would instead be making $41,400 if inequality had not changed since 1980.
In other words, if growth had been evenly shared, this household would have earned 36% more.

Belief Elicitation (Left) and Information Treatment (Right) from Hvidberg, Kreiner & Stantcheva (2022)

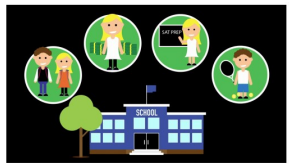




Example of a narrative information treatment: Explaining some of the longstanding causes of racial gaps from Alesina, Ferroni, and Stantcheva (2021)

(A)

(B)

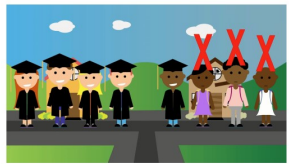


(C)

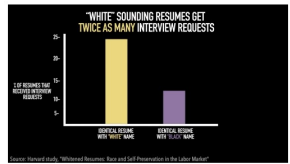
(D)

(E)

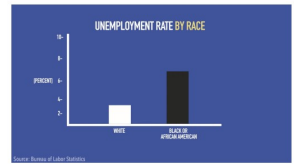
(F)



(G)



(H)



Example of a narrative information treatment: Description of a day in the life of a hard-working immigrant from Alesina, Miano, and Stantcheva (2016)

Emma legally came to the U.S. at age 25.

She lives with her husband - a construction worker - and two small children in a one-bedroom apartment.

For the past 5 years, she has been working in a retail store.

She starts work at 5 am every day of the week, earning the minimum wage for such tasks as restocking the shelves, helping customers, mopping the floor and cleaning the bathrooms.

When her day shift at the store ends at 3 pm, Emma starts her second job as a cleaning lady.

She takes two buses to get to her clients.

She finishes around 7 pm and gets home by 8 pm.

She then makes dinner for her family and sometimes helps the children with their homework before they go to bed.

Emma takes online courses. She stays up until midnight to work on her courses.

She cannot take out a loan to go to a full-time college.

Emma and her husband have no free time, no weekends, and haven't taken any holidays since arriving in the U.S..

Despite working two jobs and barely making ends meet, Emma is very happy to be in the U.S..

She hopes that thanks to her hard work she will one day be able to start her own small business.

Factorial experiments: experimentally varying attributes in hypothetical situations

Vignettes: short descriptions or stories that vary across experimental conditions along key factors of interests.

Conjoint designs: often refer to tables or list descriptions of people/situations that only show attributes and level. More direct, can make specific features salient.

Both can have simple design (only one profile) or paired design (presenting profiles that respondent rates or ranks)

Benefits: realistic scenarios that manipulate effects of interest. Allow for more complexity and testing multiple hypotheses at once. May limit liklihood of social desirability bias.

Challenges: external validity. Would people act the same way in “real life"?

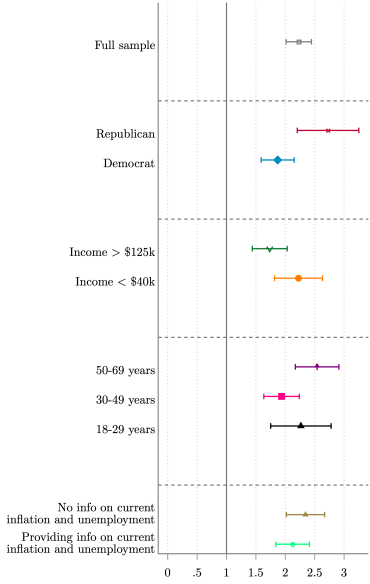
Conjoint from Binetti et al. (2024)

If you had to pick, which of the following scenarios would you prefer to live in for the next year in the US?

Scenario 1	Scenario 2
○	○

	Scenario 1	**Scenario 2**
Unemployment	6%	10%
Inflation	12%	8%

Conjoint experiment: Implied weight on inflation relative to unemployment



Paired Conjoint from Hainmueller et al. (2015)

Please take a thorough look at the two applicant profiles and then make your decision.

	Applicant 1	Applicant 2
Sex	Female	Male
Country of Origin	the Netherlands	Bosnia and Herzegovina
Age	55 years	55 years
In Switzerland since	Birth	14 years
Educational Status	Mandatory schooling	Mandatory schooling
Language Proficiency	Speaks German fluently without accent	Speaks accent-free Swiss German
Integration Status	Hardly different from a Swiss	Is integrated into Switzerland

Do you want the applicants to be granted Swiss citizenship?

	Yes	No
Applicant 1	☐	☐
Applicant 2	☐	☐

Design issues and practical recommendations for factorial experiments

Design issues

- Table formats help: can randomize in a way not possible with texts, clearer and less tiring to read
- Not advisable to have pure between-person design
- Mixed designs: different groups of respondents see different groups of vignettes. So, you can make comparisons both between and within respondents.

Practical recommendations

- Avoid implausible combinations and keep cognitive load manageable
- Randomize conditions and attribute order
- Improve level of immersion by using appropriate media
- Choose the right attributes and settings.
- Choose right causal identification: decomposition e.g., ANOVA or multilevel modeling. Average marginal component effect (ACME) often used.

Follow-up surveys and persistence

Concern: treatment effects are temporary. Especially concerning if estimated initial impact is due to SDB or EDE. Can also dissipate naturally, e.g. if treatment on topics frequent in daily life or treatment isn't salient/interesting.

Degree of persistence depends on type. For example, priming treatments' effects likely to dissapate quickly. Persistence most relevant for information or pedagogical

Recontact rates differ across survey and time between first survey and follow-up. To maximize recontact, think about increasing incentives and make it as easy as possible.

Design

- Questions must be identical.
- Be sure to address differential attrition appropriately.

Conclusion

Surveys are a unique opportunity to dive into people's minds to better understand how they reason, the things they care about, and their preferences

For high quality data and reliable results, key to focus on design, sampling, and analysis

More in the following two papers (find both on the Social Economics Lab website)

"How to Run Surveys: A guide to creating your own identifying variation and revealing the invisible" by Stefanie Stantcheva

"'Understanding Economic Behavior Using Open-ended Survey Data," by Ingar Haaland, Christopher Roth, Johannes Wohlfart, and Stefanie Stantcheva