Designing Information Provision Experiments

Christopher Roth[1]

[1]Warwick, CAGE, CESifo, CEPR, briq

PhD course on Subjective Beliefs in Macroeconomics and Household Finance, April 19, 2021

Motivation I

- **Expectations** play a central role in **macroeconomics** and (household) **finance**.
- Therefore it is important to understand...
  - ... how households and firms **form** expectations.
  - ... how expectations **causally** affect household and firm behavior.
- To shed light on these questions we need a **toolkit** to...
  - measure **expectations** about and **attention** to economic variables.
  - **exogenously vary expectations and attention** to provide causal evidence.

Motivation II

- Studies on expectations are becoming increasingly common in macro and finance.
  - Increasing skepticism towards rational expectations.
  - Increased focus on non-standard expectation formation.
- This is part of a wider trend to use micro data in macro.
- Trend to use survey data to microfound macro-issues.

Motivation III

Narayana Kocherlakota, former Minnesota, freshwater economist extraordinaire, in a short note "Thoughts on 'The Trouble with Macroeconomics"' (2016):

> *We need to encourage those who are trying to learn more about how people actually form expectations. [...] At the same time, we need to be a lot more flexible in our thinking about models and theory, so that they can be firmly grounded in this improved empirical understanding.*

Goal of this lecture

- Provide you with an overview of **state-of-the-art methods** to...
    - measure beliefs and expectations
    - design information interventions
    - design outcomes to minimize the relevance of anchoring and demand effects.

Reference

These slides are based on a review article.

Haaland, Roth and Wohlfart (2021). “Designing Information Provision Experiments”

The review provides in-depth coverage of design issues

`https://osf.io/kcj7g/`

Outline of talk

Measuring Beliefs

- Standard elicitation techniques: the toolkit
- Framing and incentives

Information treatments

- Attention versus information
- Mitigating anchoring and demand effects
- Practical advice
- Connecting experiments with theory

Surveys – Agenda

Measuring Beliefs

Standard elicitation techniques: the toolkit
Framing and incentives

Information treatments

Attention versus information
Mitigating anchoring and demand effects
Practical advice
Connecting experiments with theory

Standard elicitation techniques: the toolkit

An example

- Example: say you are interested in respondents' beliefs about the state of the Dow Jones in one year.
- How would you measure that belief?
- What considerations would you have on your mind when deciding on the right measure?

Design considerations

- Who are the participants?

- How much heterogeneity in cognitive abilities will there be in your sample?

- Precise questions are great as they are conducive to increasing **interpersonal comparability** and as they **map more closely to models**.

- What **models** do you want to speak to?

- Is your research question focused on **measurement** or **causal effects**?

Four types of broad elicitation techniques

- Qualitative questions (likert scale)
- Quantitative point beliefs
- Probabilistic elicitations
- Qualitative open-ended questions

Qualitative beliefs I

- Example: say you are interested in respondents' beliefs about the state of the Dow Jones in one year.
- You could ask people a simple qualitative question: how likely is it that the Dow Jones will be higher than today in 12 months from now? (Very unlikely, unlikely, likely, very likely)

Qualitative beliefs II

To what extent do you agree with the following statements?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The US stock market will have high returns over the next ten years.	○	○	○	○	○
The total net wealth of my household will increase considerably over the next ten years.	○	○	○	○	○

Next >>

What are the **advantages/disadvantages** of this approach?

Qualitative beliefs III

- **Advantage**: the response options do not require any quantitative skills and therefore should be easy to understand.

- **Disadvantages**:

  1. Response options are not easily interpersonally comparable.
  2. People might hold systematically different beliefs about what the question means.
  3. Verbal response scales are relatively crude and therefore limit the extent of information that can be conveyed.

Quantitative beliefs I

- Respondents are asked to state their beliefs on **a numerical scale**.
- E.g. At what level will the Dow Jones be in 12 months from now?
- It's good practice to ask a qualitative question about confidence in beliefs. E.g. How confident are you in your previous estimate? (Very confident, confident, not confident, not confident at all)

Quantitative beliefs II

Now we would like to ask you about your views on the development of different economic indicators over the next ten years.
Over the next ten years, what do you think will be...

the average annual change in home prices: ______ percent.

the average annual growth rate of real US GDP: ______ percent.

the average annual inflation rate: ______ percent.

the average annual change in your total household labor income: ______ percent.

the average annual change in rent on homes/apartments: ______ percent.

the average annual return of the US stock market: ______ percent.

What are the **advantages/disadvantages** of this approach?

Quantitative beliefs III

- Advantage: Interpersonal comparability and relatively straightforward for respondents to understand.

- Clear disadvantages:
  - Do not allow for individuals to express their **uncertainty** about outcomes.
  - It is unclear which feature of their subjective belief distribution over potential future outcomes respondents report.
    - While researchers often implicitly or explicitly interpret point beliefs as the mean over the respondent's subjective distribution, **respondents may report their median or mode belief**.

Probabilistic beliefs I

- In probabilistic belief elicitation, respondents **state probabilities for the occurrence of different mutually exclusive events**.
- You elicit the entire distribution, not just a point estimate
- For example: Partition the possible values into bins, e.g. Dow Jones decreases between 0 and 5 percent, Dow Jones increases between 0 and 5 percent, Dow Jones, increases between, 5 and 10 percent, etc.

Probabilistic beliefs II: example

In this question we present you with eight possible scenarios for the **average annual return of the US stock market, over the next ten years**.
Please let us know how likely you think it is that each scenario will occur.
Please type in the number to indicate the probability, in percent, that you attach to each scenario. The probabilities of the eight scenarios have to sum up to 100 percent.
The average annual US stock market return **over the next ten years** will be…

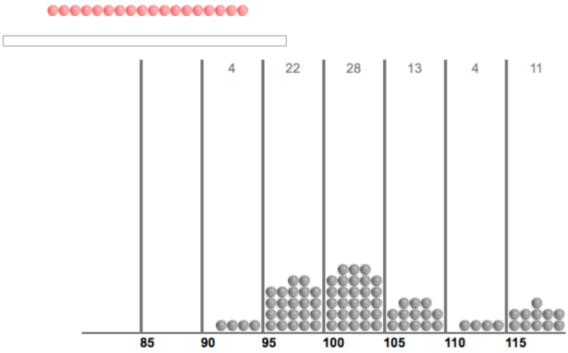
Scenario 1: … more than 20 percent:	0	%
Scenario 2: … between 10 and 20 percent	0	%
Scenario 3: … between 5 and 10 percent	0	%
Scenario 4: … between 0 and 5 percent	0	%
Scenario 5: … between -5 and 0 percent	0	%
Scenario 6: … between -10 and -5 percent	0	%
Scenario 7: … between -20 and -10 percent	0	%
Scenario 8: … less than -20 percent	0	%
Total	0	%

Next >>

Probabilistic beliefs III: increasing comprehension

- Increasing comprehension of elicitation
- Then endow respondent with 100 balls / points that they can allocate to different bins to express their relative confidence
- Ideally use a **visual tool** to make the procedure more intuitive and graphically represent pdf

Probabilistic beliefs: Visual tool



Common problems I

- Problem with eliciting beliefs in surveys:
- We model people's beliefs as well-defined subjective probability distributions
- But **many people do not hold well-defined probability distributions**!
  - Focal point responses (50:50)
  - Internally inconsistent and highly volatile answers

Common problems II

- Ket take-aways for survey design:
  - **Not overburdening** participants with excessively complex questions
  - ... even if this comes at the cost of **sacrificing some "rigor"** from the viewpoint of economic models
  - Using **intuitive elicitation formats**
  - Assessing the extent to which responses reflect **genuine beliefs** rather than confusion

Which beliefs should you measure?

- When is it ok to use simple measurement, and when does one need probabilistic beliefs?
- All depends on your research questions. E.g. does the model you want to test feature an important role for uncertainty?
- Are you interested in quantification of treatment effects?
- Is your main object of interest a behavioral outcome and the beliefs are just a manipulation check?

Bin effects I

- Related issue:
- We know from lots of work in experiments that there are **strong bin / partition effects** that have big impacts on the beliefs people state
- Example:
  - **Version A**: What is the probability that 10 years from now you will:
    - Work full-time (at least 40 hours per week)
    - Work less than full-time
  - **Version B**: What is the probability that 10 years from now you will:
    - Work full-time (at least 40 hours per week)
    - Work between 35 and 40 hours per week
    - Work between 30 and 35 hours per week
    - Work between 25 and 30 hours per week
    - Work between 20 and 25 hours per week
    - Etc.

Bin effects II

- You see where this is going: people state lower beliefs for working full-time in version B
- Need to be aware of this in designing surveys

Open-ended responses: Unstructured text data

- It is also possible to use **open-ended questions** to measure beliefs – in an open text box (Andre et al., 2021; Bursztyn et al., 2020; Stantcheva, 2020).
  - Text as data (Gentzkow et al, 2018).

- The key advantage of such open-ended questions is that respondents **are not primed** by the available answer categories.

- The combination of priming and open-ended questions, allows one to shed light on which **assocations** are the drivers of the effects induced by priming (Andre et al., 2021)

Open-ended responses: an example

Your thoughts

Above, you predict how the change in the alternative scenario affects the US economy.
Please tell us how you come up with your predictions.
What are your main considerations in making those predictions?
Please respond in 2-3 sentences.

→

Questions?

Belief Elicitation: Details

Leveraging external benchmarks in belief elicitations

- Measuring subjective beliefs about quantities with **well-defined benchmarks** (e.g. current u-rate) has several advantages.
    1. The experimenter fixes beliefs about the environment and imposes additional structure on the responses.
        - This in turn may lower **heterogeneity in how the question is interpreted**.
    2. It allows one to characterize the extent of **biases in beliefs** compared to the benchmark.
    3. It enables one to **incentivize** the belief elicitations in a transparent way.

Framing of belief elicitations I

- In settings in which respondents are relatively **experienced**, they are capable of accurately assessing economic quantities.
  - For example, respondents are relatively good at assessing the **price of gas** (Ansolabehere et al., 2013).

- However, in settings in which respondents are relatively unfamiliar, there will be higher levels of measurement error.
  - This is especially true when respondents are **unsure about the response scale**, such as in the context of inflation expectations.

Framing of belief elicitations II

- However, careful framing of questions can **reduce measurement error**.

- For instance, Roth et al. (2019) measure beliefs about the debt-to-GDP ratio in the US using different historical or cross-country anchors.
    - Show that the provision of an anchor reduces the dispersion of beliefs and rounding.

Incentives

- Incentives have been shown to **reduce partisan bias** in people's stated beliefs (Bullock et al., 2015; Prior et al., 2015).
  - The partisan gap in **beliefs about the current unemployment rate shrinks** when respondents receive prediction incentives.
- In the context of macroeconomic forecasting, it has been shown that unincentivized survey reports **strongly correlate** with incentivized belief measures (Armantier et al., 2015).
- **Prediction incentives do not** have any statistically significant effects on reported beliefs about macro-variables (Roth and Wohlfart, 2020).
- However, in the context of **judgment tasks** higher incentives do improve performance (Camerer and Hogarth, 1999).

Questions?

Outline of talk

Measuring Beliefs

- Standard elicitation techniques: the toolkit
- Framing and incentives

Information treatments

- Attention versus information
- Mitigating anchoring and demand effects
- Practical advice
- Connecting experiments with theory

Information treatments

Information provision experiments I

- Correlation between beliefs and behavior is confounded for several reasons:
  - **Reverse causality** (e.g. induced by motivated beliefs)
    - E.g. people who just bought a house want to believe that house prices will further increase.
  - **Omitted variable bias** (e.g. character traits)
    - E.g. people with optimistic personality traits have both optimistic beliefs about future income and a low savings rate.
  - **Measurement error** in beliefs
    - People make errors in probabilistic belief elicitations.

Information provision experiments II

- To get **causal estimates** of beliefs on behavior researchers provide respondents with information.

- Standard design proceeds as follows:

  1. Measure **prior beliefs** about the variable of interest (e.g. likelihood of a recession in 2022).
  2. Researchers provide treatment group with information (e.g. forecast about likelihood of a recession in 2022 from a professional forecaster) and a control group with no information
  3. Measure behavior of interest (e.g. consumption behavior)
  4. Measure post-treatment beliefs (e.g. personal income expectations)

Two types of papers

There are broadly speaking two types of papers in this literature

1. What is the **causal effect of expectation Y on behavior X** ?
2. **How do people form their expectations** about variable Y?

Causal effect of expectation Y on behavior X

1. What is the causal effect of inflation expectations on firm decisions (Coibion et al., 2019b)?
2. What is the causal effect of recession expectations on households' consumption decisions (Roth and Wohlfart, 2020)?

How do people form their expectations about variable Y

1. How do people's home price growth expectations adjust in response to information about past home price growth (Armona et al., 2019)?
   1.1 If households upward adjust their future houseprice expectations in response to an upward adjustment of perceived past home price growth → belief in **momentum**.
   1.2 If households downwards adjust their future houseprice expectations in response to an upward adjustment of perceived past home price growth → belief in **mean reversion**.
2. How strongly do people update their inflation expectations in response to information about past inflation in **volatile vs stable inflation environments** (Cavallo et al., 2016)?
   2.1 Strong adjustments in stable inflation environments compared to small adjustments in volatile environments, consistent with **rational inattention**.

Questions?

Attention versus information

Disentangling information from attention

- One key challenge in information experiments is to disentangle the effects of **priming/attention** from **genuine belief updating**.

- Common methods to **mitigate concerns about priming** include

  1. eliciting **prior beliefs** of respondents in both the treatment and the control group
  2. separate the information provision from the main outcomes with **follow-up studies**, and
  3. to include an **active control group** (that is, the control group also receives (differential) information).

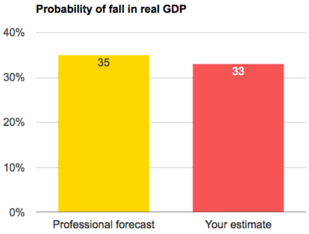
Active control group: An example

Information about the likelihood of a recession:

You said that you think that the probability of a fall in real US GDP in the fourth quarter of 2017 is **33 percent**.

We now would like to provide you with information on the view of a **professional forecaster** on the likelihood of a recession.

According to a financial services provider that regularly takes part in a survey of professional forecasters by the Federal Reserve Bank of Philadelphia, the probability of a fall of real GDP in the fourth quarter of 2017 is **35 percent**.



Advantages of active control group designs

- In a design with a pure control group the variation hinges on prior beliefs. The identification mostly comes from **individuals with larger misperceptions** ex ante.
  - An active control group design generates variation in the relevant belief also among individuals with more accurate priors.
- Receiving an information treatment may have side effects, such as **uncertainty reduction, attention, and emotional responses**.
  - Such side effects should arguably be constant across groups that receive different pieces of information.
- Prior beliefs are **measured with error** and correlated with unobservables.
  - Thus, **causal identification** and the interpretation of heterogeneous treatment effects are more difficult in pure control designs.

Advantages of pure control designs

- Having a pure control group makes it easier to interpret correlations between the pre-treatment beliefs and the outcome of interest.
- Sometimes the **policy relevant question of interest** is concerned with the effect of providing a particular piece of information compared to not providing this information.
  - How do people change their inflation expectations when they hear about central bank communication (Coibion et al., 2020)?
- Sometimes it is not possible to have an active control group without **deceiving** respondents

Cross-learning

- Respondents may not only update beliefs about the object of interest, but at the same time change their beliefs about other variables.
    - Coibion et al. (2019a) find that provision of information about inflation not only changes respondents' inflation expectations but also their beliefs about GDP growth.

- Cross-learning can complicate the interpretation of instrumental variables (IV) estimates exploiting randomized information provision.

- In the presence of substantial cross-learning it is **less straightforward to interpret the effects of information on behavior through the lens of belief changes**.

Dealing with cross-learning

- One way to over-come the issue of cross-learning is to **hold fixed beliefs about other variables by providing the same information about other variables** to respondents in both the control and the treatment groups.
  - However, simultaneous provision of several pieces of information might arguably reduce attention to the main piece of information and lead to a weaker first stage.

- Researchers should **include measures for beliefs about other variables** which could be shifted by the treatment in their survey in order to be able to detect cross-learning.

Key design issues

- What **type of information** to provide?
  Quantitative/qualitative/probabilistic.

- What is the **source of the information**?

- What is the **identity** of the sender of the information?

Questions?

Mitigating anchoring and demand effects

Mitigating anchoring and demand effects I

- Information treatments are an important tool to get at **causality**.
- However, they may have **undesirable side effects**: demand and anchoring effects.
- The next few slides are about the **mitigation** of these undesirable side effects.

Mitigating anchoring and demand effects II

- A concern for quantitative outcome measures elicited after the information provision, such as posterior beliefs about the statistic, is unconscious numerical anchoring.

- Best practices:

  1. Measuring a quantitative beliefs on a **scale that differs from the scale on which the information is communicated**.
  2. One should also employ **qualitative measures of beliefs**, which are naturally immune to numerical anchoring.
  3. **Follow-up surveys**, conducted a few weeks after the initial information intervention, are an important tool used to mitigate concerns about numerical anchoring.
     - Numerical anchoring is a **short-lived phenomenon**.

Mitigating anchoring and demand effects III

- **Obfuscation in experiments**
  - Hiding the purpose of the experiment.
- Obfuscated follow-ups (Haaland and Roth, 2020, 2021)
- Obfuscated information treatment
  - Giving a cover story for the treatment information

Mitigating anchoring and demand effects IV

- **Obfuscated follow-ups** (Haaland and Roth, 2020, 2021) adjusted to a macro application

- **Only administer the information treatment in baseline survey** and do not collect any of the main outcome variables.
  - E.g. give different professional forecasts about the future unemployment rate.

- Survey company reinvites respondents a few weeks later to a **seemingly unrelated survey**, in which the main outcomes (e.g. consumption behavior in the last week) is collected.
  - Use **different survey layouts**
  - Mention only the affiliation of a subset of different researchers involved in each wave (study from Uni Bergen vs. study from U Warwick).
  - Ask a series of **unrelated questions**.

Mitigating anchoring and demand effects V

- **Admin data** (Bottan & Perez-Truglia, 2020; Laudenbach, Weber and Wohlfart, 2021)
- **Incentivized measures embedded in survey**
  - Donations to NGO
- Make outcome measures **anonymous**
  - Anonymous online petitions (Grigorieff et al. 2020)
  - List method or randomized response technique

Mitigating anchoring and demand effects VI

- Measuring **beliefs about the experimenter expectations/study purpose**.
  - "What do you think is the percent chance that the experimenters expected you to choose action X"?
  - Moreover, researchers can use open-ended questions to measure beliefs about the purpose of the study (Bursztyn et al., 2020).
- **"Demand treatments"** in order to assess how sensitive behavior in a given setting is to signals about the experimenters' expectations (de Quidt et al., 2018).
  - Roth & Wohlfart (2020) find little responsiveness to demand treatments in the context of macro-expectations.

Questions?

Practical advice

Measuring attention in online surveys

- Online surveys are a popular and, by now, cheap tool to conduct surveys.
- Inattention in online surveys is very high.
- Measuring inattention is key.
  - Use standard screener questions (Berinsky, 2014)
  - Use open-ended questions
- Screen out bots using CAPTCHAS and open-ended questions.

Measuring attention in online surveys: An example

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. **To show that you read our questions carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer in the below question**

Given the above, how interested are you in politics?

- [ ] Very strongly interested
- [ ] Very interested
- [ ] A little bit interested
- [ ] Almost not interested
- [ ] Not at all interested

→

Which online samples are out there?

Different types of samples:

- **Probability-based** samples
- Online panels **representative in terms of observables**
- **Online labor markets**, such as Amazon Mechanical Turk

Probability-based Samples

- The most representative samples are probability-based panels.
- In a probability-based panel, the survey company recruits the sample by randomly selecting households from a representative sample frame.
- People cannot join the panel unless they have been randomly selected for participation.
- **Advantage**: Useful for descriptive papers whose objective of interest is the estimation of a population mean (and margin of error)
- **Disadvantage**:
  - Quite costly
  - Still very low response rate (between 5% and 15%).

Online panels representative in terms of observables

- Representative online panels are constructed to be representative of the general population **in terms of observable characteristics**.
- The survey company recruits respondents through advertisements and **anyone who wants to join can do so**.
- The main advantage of these panels is that they are quite affordable (2$ per participants for a 10 minute survey) while retaining **representatives** in terms of important **observables**.
- **Drawbacks:**
  - Lack of random sampling makes it difficult to estimate the margin of error for the general population.
  - These samples do not include respondents from the non-Internet population.

What makes a good (information provision) experiment?

Just ask yourself the following questions:

- How does your experiment inform **theoretical models**?
- How does your experiment inform **policy debates**?
- How **natural** is your outcome measure? Is it **similar** to decision people take in **the real world**? Are these decisions **important**?
- Does your treatment affect beliefs about an object of interest that plays an important role in models or in the public debate?

Connecting experiments with theory

- The most powerful experiments usually have **sharp implications for different theories** of economic behavior or theories of expectation formation.
- Testing the **key behavioral assumptions** of classes of models.
- **Calibrating models** using experimentally estimated expectations adjustments (Roth, Wiederholt, Wohlfart, 2021)

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective Models of the Macroeconomy: Evidence from Experts and a Representative Sample," *Available at SSRN 3355356*, 2021.

Ansolabehere, Stephen, Marc Meredith, and Erik Snowberg, "Asking about numbers: Why and how," *Political Analysis*, 2013, *21* (1), 48–69.

Armantier, Olivier, Wändi Bruine de Bruin, Giorgio Topa, Wilbert Klaauw, and Basit Zafar, "Inflation Expectations and Behavior: Do Survey Respondents Act on their Beliefs?," *International Economic Review*, 2015, *56* (2), 505–536.

Armona, Luis C, Andreas Fuster, and Basit Zafar, "Home Price Expectations and Behavior: Evidence from a Randomized Information Experiment," *Reviw of Economic Studies*, 2019.

Bullock, John G, Alan S Gerber, Seth J Hill, and Gregory A Huber, "Partisan bias in factual beliefs about politics," *Quarterly Journal of Political Science*, 2015, *10* (4).

Bursztyn, Leonardo, Ingar Haaland, Aakaash Rao, and Christopher Roth, "Disguising Prejudice: Popular Rationales as Excuses for Intolerant Expression," NBER Working Paper 27288 May 2020.

Camerer, Colin F and Robin M Hogarth, "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of risk and uncertainty*, 1999, *19* (1-3), 7–42.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia, "Learning from potentially-biased statistics: Household inflation perceptions and expectations in Argentina," *Brookings Papers on Economic Activity*, 2016, pp. 59–108.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij, "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial," *National Bureau of Economic Research Working Paper*, 2019.

_, Yuriy Gorodnichenko, and Tiziano Ropele, "Inflation expectations and firm decisions: New causal evidence," *Quarterly Journal of Economics*, 2019.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Prior, Markus, Gaurav Sood, Kabir Khanna et al., "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions," *Quarterly Journal of Political Science*, 2015, *10* (4), 489–518.

Roth, Christopher and Johannes Wohlfart, "How Do Expectations About the Aggregate Economy Affect Personal Expectations and Behavior? Experimental Evidence," *Review of Economics and Statistics*, 2020.

__ **, Sonja Settele, and Johannes Wohlfart**, "Beliefs about Public Debt and the Demand for Government Spending," *Journal of Econometrics*, 2019.

Stantcheva, Stefanie, "Understanding Economic Policies: What do People Know and How Can they Learn?," Technical Report, Harvard University Working Paper 2020.