ANNUAL REVIEWS

Annual Review of Economics

How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible

Stefanie Stantcheva[1,2,3]

[1]Department of Economics, Harvard University, Cambridge, Massachusetts, USA; email: sstantcheva@fas.harvard.edu

[2]National Bureau of Economic Research, Cambridge, Massachusetts, USA

[3]Centre for Economic Policy Research, London, United Kingdom

Annu. Rev. Econ. 2023. 15:205–34

First published as a Review in Advance on March 27, 2023

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-091622-010157

JEL codes: C8, C9, A22, A23, D7, P16

Keywords

surveys, experiments, priming, information, beliefs, perceptions, sampling

Abstract

Surveys are an essential approach for eliciting otherwise invisible factors such as perceptions, knowledge and beliefs, attitudes, and reasoning. These factors are critical determinants of social, economic, and political outcomes. Surveys are not merely a research tool. They are also not only a way of collecting data. Instead, they involve creating the process that will generate the data. This allows the researcher to create their own identifying and controlled variation. Thanks to the rise of mobile technologies and platforms, surveys offer valuable opportunities either to study broadly representative samples or to focus on specific groups. This article offers guidance on the complete survey process, from the design of the questions and experiments to the recruitment of respondents and the collection of data to the analysis of survey responses. It covers issues related to the sampling process, selection and attrition, attention and carelessness, survey question design and measurement, response biases, and survey experiments.

1. INTRODUCTION

Surveys are an invaluable research method. Historically, they have been used to measure important variables, such as unemployment, income, or family composition. Today, often there are high-quality administrative or other big data that we can use for this purpose. However, some things remain invisible in nonsurvey data: perceptions, knowledge and beliefs, attitudes, and reasoning. These invisible elements are critical determinants of social, economic, and political outcomes.

As economists, we typically tend to prefer revealed preference approaches, which involve inferring unobserved components from observed behaviors and constraints. These methods are useful and suitable for a wide range of questions. However, when it comes to measuring and identifying the above-mentioned invisible factors, there are many challenges. One could, in principle, specify a complete structural model of beliefs or other invisible factors and use observational or quasi-experimental data on some behaviors to estimate these underlying factors. However, this requires many assumptions and identifying variations that may be absent in the data. For instance, suppose you wanted to measure people's beliefs about whether a carbon tax would reduce car emissions or whether trade policy will lead to adverse distributional consequences. You would likely have difficulty finding behaviors that allow you to identify these perceptions. There are plenty of other examples of perceptions, beliefs, attitudes, or reasonings that profoundly shape our views on policy and social issues but that we do not necessarily reveal with our microeconomic, observed behavior. Surveys are an essential approach for eliciting these intangibles more directly.

Surveys are not merely a research tool—they are part of a unique and distinct research process. Furthermore, surveys are not only a way of collecting data. Unlike when using observational data, you as the researcher are the one creating the process that will generate the data. You can therefore create your own controlled and identifying variation. This process presents many opportunities as well as challenges. Your survey design is an integral part of your research process.

Thanks to the rise of mobile technologies and platforms, online surveys offer valuable opportunities either to study broadly representative samples or to focus on specific groups. They are flexible and customizable and can be made appealing and interactive for respondents. They allow researchers to conduct large-scale investigations quickly—sometimes in real time—and explore new questions. They are indeed a way to engage with people and get a glimpse of their mental processes.

In order to use surveys for economic research in a fruitful way, there are, however, important issues to take into account. This article provides a practical and complete guide to the whole survey process, from the design of the questions and experiments to the collection of data and recruitment of respondents to the analysis of the survey responses. The goal is to give researchers from many fields and areas practical guidance on leveraging surveys to collect new and valuable data and to answer original questions that are challenging to answer with other methods. The examples used in the article come from a wide range of fields—a testimony of the extensive use of survey methods. Although the article focuses on written surveys, specifically online ones, many of the concepts and tips apply to surveys more broadly, regardless of the mode.

When you decide to run a survey, you may wish to start writing the questions quickly. However, do not jump into this before a lot of careful thinking. There may be a temptation to think about writing your survey as just the equivalent of getting the data in observational empirical work. However, you are the one creating the data here, which gives you many opportunities and presents many challenges. Writing your survey questions is already part of the analysis stage.

You first need to outline very clearly what your research question is. There is no such thing as a good survey or a good question in an absolute sense (although there are bad surveys and bad questions). A good survey is adapted to your research issue. Therefore, when writing survey questions, you must always remember how you will analyze them; the right design will depend on your goal.

Well-designed survey questions allow you to create your own controlled variation. This distinguishes social and economic surveys from other types of surveys. The goal is not only to collect statistics but also to understand reasoning, attitudes, and views and to tease out relationships. When you design your questions, you need to keep the concept of ceteris paribus, or "all else equal," in mind and think of the exercise as creating your own controlled (identifying) variation. Each question needs to ask about only one thing at a time and hold everything else as constant as possible (and respondents need to be aware of that).

Supplemental Material >

In **Supplemental Appendix A-3**, I outline some best practices for writing questions, based on the many references cited in this review and my own experience. **Supplemental Appendix A-3.3** builds extensively on the work of Dillman et al. (2014) and Pew Res. Cent. (2022). Some of the examples there are intentionally adapted to be more suitable for economic surveys, with a few examples used with very minor modifications.

The article is organized as follows. Section 2 describes the sampling process and recruiting of respondents. It also discusses how to deal with selection and differential attrition. **Supplemental Appendix A-1.5** contains more detailed explanations of procedures to correct for selection and attrition. Section 3 discusses methods for detecting, minimizing, and dealing with different response biases that can arise in surveys. It covers biases related to the choice of answer options unrelated to their content (moderacy bias, extreme response bias, and ordering bias), acquiescence bias, experimenter demand effect (EDE), and social desirability bias (SDB). Section 4 offers guidance on conducting survey experiments. The online **Supplemental Appendix** also contains useful supplementary materials, including reviews of many papers relevant to each section in the main text.

In addition, due to space constraints, two important issues are only covered in the online **Supplemental Appendix**. First, once you have recruited a high-quality sample, the essential asset in your survey is your respondents' attention. As is the case for many other survey issues, the condicio sine qua non in dealing with respondents' attention or lack thereof is a good survey design, as covered in **Supplemental Appendix A-3**. Beyond that, there are some targeted methods. **Supplemental Appendix A-2** presents methods to foster respondents' attention and minimize careless answers, as well as methods to screen for inattentive respondents or careless answer patterns. Second, **Supplemental Appendix A-3** dives into the design of survey questions. It covers general best practices, open-ended questions, closed-ended questions, visual design, measurement issues, monetary incentives and real-stakes questions, and the ordering of questions.

2. SAMPLE

2.1. Types of Samples

The first question is what kind of sample you need for your research question. A nationally representative sample can be valuable in many settings, while a more targeted sample—e.g., one obtained by oversampling minorities or specific age groups, or restricting to employees or job seekers, etc.—may be more appropriate in others.[1] A useful notion is "sampling for range" (Small 2009)—i.e., the idea that your sample should be intentionally diverse in terms of conceptually important variables. **Supplemental Appendix A-1.1** reviews different sampling methods.

[1]Surveys can also be done for firms, instead of individuals or households, as done by Link et al. (2022) and Weber et al. (2022).

There are different types of survey channels you could use to build your sample, listed below.

- Nationally representative panels—e.g., in the United States, the Knowledge Panel,[2] NORC's AmeriSpeak,[3] and the Understanding America Study.[4]
- Commercial survey companies that use quota sampled panels, such as Qualtrics, Dynata, Bilendi, and Prolific Academic.
- Commercial survey marketplaces (such as Lucid), which are very similar to commercial survey companies but require more in-house and hands-on management of the survey process by the researcher. I will discuss these together with the survey companies.
- Convenience samples, which, as the name indicates, are sample populations that are convenient for the researcher to access. Examples are university students or conference participants.
- Online work platforms like Amazon's Mechanical Turk (MTurk), which are in between convenience samples and quota sampled panels, given the large pool of respondents.
- Targeted groups from specific pools, such as experts, employees at a firm, economists, etc.
- Governments' or institutions' surveys—e.g., surveys run by Statistics Denmark[5] for matching tax data with survey data, or the Survey of Consumer Expectations.[6]

These survey channels differ in the control they give you over the recruiting process of your respondents. Therefore, the advice below tries to distinguish between doing a survey in-house, with complete control over your process, and using a platform with a given process in place. **Supplemental Appendix A-1.2** provides information pooled from several survey companies' documentation about their recruitment processes, rewards, and pools of respondents.

Supplemental Material >

Sometimes you may be able to use mixed-method surveys to reach different types of respondents (e.g., online plus phone survey or online plus door-to-door surveys). For instance, the Understanding America Study recruits panel members through address-based sampling with paper invitation letters. Individuals who lack Internet access are provided with a tablet and an Internet connection, which increases the coverage rate of this panel. While mixed methods could introduce discrepancies between respondents who answer through different modes, they could be particularly valuable if you are interested in surveying populations that are less likely to be online, including segments of the population in developing countries. This article's content and visual design issues also apply to mixed methods. However, there are specificities to consider if the survey is over the phone or in person.

2.2. Survey Errors and Selection into Online Surveys

This section discusses how online survey respondents compare to the target populations, starting with a review of survey errors.

2.2.1. Survey errors and threats to representativeness. **Figure 1** illustrates the different stages of the sampling process, from the target population to the final sample, and the errors that occur at each stage. The target population is your population of interest, for example, all adults 18–64 in the United States. The sampling frame or pool of potential respondents represents all

[2] **https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel**.
[3] **https://amerispeak.norc.org/us/en/amerispeak/about-amerispeak/overview.html**.
[4] **https://uasdata.usc.edu/index.php**.
[5] **https://www.dst.dk/en.**
[6] **https://www.newyorkfed.org/microeconomics/sce#/**.

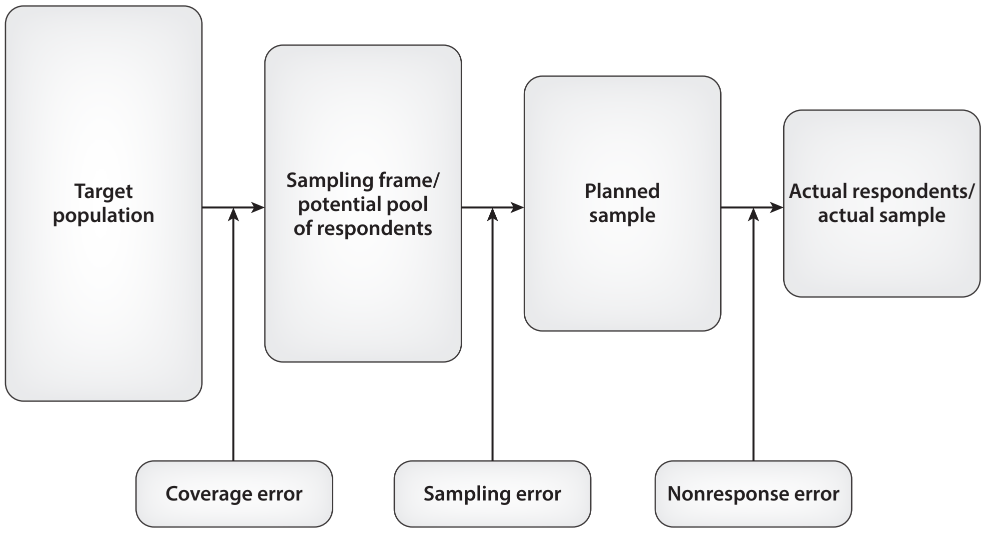

Figure 1

From the target population to the actual sample.

the people in the population you can potentially sample and invite to the survey. One bias occurs from coverage error, which is the difference between the potential pool of respondents and the target population. For instance, in online surveys, you will not be able to sample people who are not online. The planned or target sample is all the people you would ideally like to complete your survey. The difference between the planned sample and the sampling frame is due to sampling error, that is, the fact that you are drawing only a sample from the full sampling frame. Different types of sampling are reviewed in **Supplemental Appendix A-1.1**. For instance, probability sampling will lead to random differences between your target sample and the sampling frame. The actual sample or set of respondents represents the people who end up taking your survey. Nonresponse error refers to the differences between the target sample and the actual sample. This error can be due to the respondent not receiving or seeing the invitation, ignoring the invitation, or following up on the invitation but refusing to participate. In general, it is difficult to distinguish between these cases (not just in the case of survey companies). Most of the time, we know little about nonresponders, other than information embedded in the sampling frame. Sometimes we do have extra information (e.g., in consecutive waves of a longitudinal survey or from additional data, such as administrative records from which the sample is selected).

Supplemental Material >

We can further distinguish between unit nonresponse bias (the difference between respondents who start the survey and those in the planned sample) and item nonresponse bias, when respondents start the survey but some answers are missing. A special case of item nonresponse is attrition, the phenomenon of respondents dropping out of the study before completing it. In this case, all items past a specific question are missing. Attrition induces a bias if it is differential, that is, not random.[7] There are ways to minimize nonresponse bias and attrition bias ex ante and ways to correct for them ex post. Conditional on respondents seeing the survey invitation, one can expect that a good design of the invitation and landing page, as explained in Section 2.3, minimizes

[7] In longitudinal surveys in which respondents are interviewed multiple times, selection into subsequent rounds is typically called attrition.

the selection of respondents based on the topic. These errors and biases will greatly depend on the survey channels and methods used. Next, we discuss the typical case of commercial survey companies.

Supplemental Material >

2.2.2. Selection in online surveys. What do we know about these survey errors in the case of commercial survey companies and survey marketplaces? **Supplemental Appendix A-1.2** provides information on their recruiting channels, processes, and pools of respondents. The sampling frame is respondents who are in the panels of the company. **Supplemental Table A-1** shows how these pools of respondents compare to the population of several countries and across two large survey companies. The sampling procedure is akin to quota sampling, which makes it difficult to estimate the sampling error and identify the planned sample. Typically, survey companies can target the invitations to background characteristics, and invitations are likely somewhat random, conditional on observed characteristics (see **Supplemental Appendix A-1**). When using survey companies, it is not easy to clearly differentiate between sampling error and nonresponse error. Because it is difficult to track the respondents in each of these stages, we can use the term "selection bias" to jointly denote the difference between respondents who start the survey and those in the target population.

Online surveys have some key advantages in terms of selection, as compared to in-person, phone, or mail surveys: (*a*) They give people the flexibility to complete the survey at their convenience, which reduces selection based on who is free to answer during regular work hours or who opens the door or picks up the phone; (*b*) the convenience of mobile technologies may entice some people who would otherwise not want to fill out questionnaires or answer questions on the phone to take surveys; (*c*) they allow surveyors to reach people who are otherwise hard to reach (e.g., younger respondents, those who often move residences, respondents in remote or rural areas, etc.); (*d*) they offer a variety of rewards for taking surveys, which can appeal to a broader group of people (especially when done through survey platforms); some rewards can appeal to higher-income respondents as well (e.g., points for travel or hotels).[8]

2.2.3. Comparing online samples to nationally representative samples. We compare the characteristics of samples from surveys using online commercial survey platforms to the characteristics of the target population across various papers in **Supplemental Appendix A-1**. **Supplemental Table A-2** shows that, in the United States, across many platforms, online samples can offer a good representation of a broad spectrum of incomes ($25,000–$100,000). However, like many other survey methods, they are not suitable for reaching the tails of the income distribution (i.e., the very poor or very rich). They tend to skew toward more educated, white (white and non-Hispanic respondents are typically oversampled whereas Black respondents tend to be

[8]While different in their goal, which is typically measurement and provision of statistics, government surveys (done over the phone, mail, or in person, now with computer-assisted technology) also face selection problems. For instance, US Census Bur. (2019) lists hard-to-survey populations, some of which could be significantly easier to reach via online surveys or other types of platforms, particularly people in physically hard-to-reach areas, dense urban areas, temporary situations (e.g., short-term renters), or younger respondents who still have mobile or Internet access. Other target populations are likely challenging to reach through any survey channel, such as people who are migrant and minorities, homeless, in disaster areas, institutionalized, seafarers and fishers, nomadic and transitory, facing language barriers, with disabilities preventing them from taking surveys, or with limited connectivity. Some other key surveys also suffer from misrepresentation of some groups, sometimes in a way that is quite different from that of online samples. For instance, Brehm (1993) shows that in the American National Election Studies and the General Social Survey, young and elderly adults, male respondents, and high-income respondents are underrepresented, while people with low education levels are overrepresented.

undersampled), and somewhat Democratic (at the expense of both Republican and Independent) respondents. Respondents from larger urban areas and urban clusters tend to be overrepresented, whereas those from medium- and small-sized urban and rural areas are often underrepresented. Some papers use online platforms to successfully replicate studies done on nationally representative or convenience samples (see Berinsky et al. 2012, Heen et al. 2020, and **Supplemental Appendix A-1.3**).

Supplemental Material >

In other high-income countries, according to **Supplemental Table A-2**, the representativity of online samples looks relatively consistent with that in the United States. However, in developing or middle-income countries, online samples are not nationally representative. Instead, they could be considered "online representative" because they represent people who are well-connected to the Internet and use mobile technologies.

Papers that match survey data to population-wide administrative data can also provide valuable information on selection into online surveys. For example, a sample recruited by Statistics Denmark looks almost identical to the target population (as in Hvidberg et al. 2021).[9]

These comparisons between the samples and the target population rely, by necessity, on observable variables. Non-probability sampling, such as the quota sampling performed by survey companies, carries risks in terms of representativeness. Therefore, it is important to always critically assess your sample in light of your survey method and topic before suggesting that your results generalize to the target population (see **Supplemental Appendix A-1.5**).

2.3. Recruiting Respondents

When using a more hands-on survey channel, you can directly control the content and format of the initial email or invitation to respondents, the number and timing of reminders, and the rewards system. On the contrary, commercial survey companies essentially handle the recruitment process (as explained in **Supplemental Appendix A-1.2**). Regardless of the survey channel used, you have complete control over your survey landing page and your survey design. You can check existing papers (including the many referenced in this article) for examples of recruiting emails and survey landing pages. It is good practice to include screenshots of your consent and landing page in your paper. If you are doing a more hands-on survey, you should also include all recruitment materials.

2.3.1. The survey landing page. The initial recruitment email and landing page of your survey are critical. You need to increase your survey engagement while avoiding selection based on your topic. Below are some general tips.

- Reduce the perceived costs of taking the survey from the start by specifying the (ideally short) survey length.
- Use simple language and a friendly visual design. Make sure everything is easily readable (on mobile devices, too), which signals to respondents that the rest of your survey will be clear and well-designed.
- Do not reveal too much about the identity of the surveyor. There is a trade-off between revealing more about yourself and your institution and telling respondents just the bare minimum for them to feel confident in taking the survey. Think about the difference between "We are a group of nonpartisan academic researchers" and "We are a group of faculty members from the Economics Department at Harvard and Princeton." On the one hand,

[9]Other papers that have matched administrative data to survey data and found good representativity include those by Karadja et al. (2017), who used paper mail surveys, and Epper et al. (2020), who invited people through paper mail to take an online survey.

revealing more may bias respondents' perceptions of the survey based on their perception of your institution (and its political leaning). On the other hand, it can provide legitimacy. Some amount of information is often required by Institutional Review Boards (IRBs), and these requirements can differ by institution. You can ask respondents whether they perceive your institution and survey as biased at the end of the study.

- Appear legitimate and trustworthy. (*a*) Think about the trade-off between revealing more about your identity and institution and revealing less. (*b*) Provide contact information so that respondents can express complaints and issues or provide other feedback. Respondents need to be able to get in touch with you. (*c*) Provide information about how the data will be stored and used. IRBs will often ask for specific language and a link to their contact and information pages. If surveys are conducted outside of the United States, there will be specific rules, such as the General Data Protection Regulation (GDPR) in the European Union. (*d*) Reassure respondents about complete anonymity and confidentiality. Survey companies have rules and agreements for respondents, but it is always good to reiterate that respondents are anonymous and their data are protected.
- Provide limited information about the purpose of the study. Some information about the survey is needed, but I would advise against revealing too much about the actual research topic to avoid selection. For instance, "This is a survey for academic research" may be sufficient, and "This is a survey for academic research in social sciences" is probably fine, too. "This is a survey for academic research on immigration," instead, will likely induce some selection based on the topic. You should never reveal the purpose or intent of the study (e.g., "We are interested in how people misperceive immigrants" or "We are interested in how information about immigrants can change people's perceptions").
- Specify some possible benefits of the survey either for research and society more broadly or for the respondent themselves (e.g., they may learn exciting things and may be able to express their opinion).
- Warn against poor response quality. If appropriate for your audience, inform respondents that careless answers may be flagged and their pay may be withheld. Note that in the case of commercial survey companies, there are typically already explicit agreements between respondents and companies on the quality of the survey responses.

Supplemental Material >

There is some trade-off between getting people interested in your topic and inducing selection bias because of it. Survey companies tend to provide little information about the survey (see **Supplemental Appendix A-1.2**). Selection is a more serious issue in some settings than others, so you must assess based on your specific situation. In surveys through commercial survey companies, I try to provide as little information as possible about the topic on the consent page (and in the first few pages of the survey). Instead, I first try to collect basic information on respondents, which will allow me to identify whether there is differential attrition or selection based on the topic. Given the large potential pool of respondents, differential attrition and selection are much greater concerns than getting a large-enough sample size. However, in another survey done on a high-quality sample with the help of Statistics Denmark (Hvidberg et al. 2021), we already have complete information on anyone in the target population and can quickly check for selection. In this case, we worry less about selection and more about maximizing engagement, since we are interested in getting a large-enough and broad sample. In such cases, the trade-off is in favor of a more informative landing page.

2.3.2. Other elements of the recruiting process. There are additional elements of the recruitment process that you will have to address unless you hire a survey company to do them for you.

- Writing an invitation email. This can be personalized to the respondent and incorporate the tips about the survey landing page discussed above.
- Sending reminders. You must plan for and send reminder emails to respondents to encourage them to take the survey.
- Ensuring that your respondents are legitimate and verified. Survey companies have several layers of verification in place (see **Supplemental Appendix A-1.2**). Following the rise in bots, automatic survey takers, and fraudsters, you will need to (*a*) employ CAPTCHAs and more sophisticated tasks at the start of the survey, such as open-ended questions (for which you can check the content) or logical questions; (*b*) not share the link publicly and only distribute it through reliable channels; and (*c*) double down on the data quality checks discussed in **Supplemental Appendix A-2**.
- Managing incentives and rewards. While survey companies will do this for you, if you are running your survey independently, you will need to set appropriate rewards and ensure you have a way to transfer rewards to respondents. Note also that, typically, respondents that are part of survey panels are, by construction, more likely to respond to surveys than those who have not signed up for surveys. If not using survey companies or panels of respondents, you will need to work hard on recruitment and incentives.
- Setting quotas. Although survey companies may do this for you, you can generally impose your quota screening at the start of the survey. This involves asking respondents some screening questions and channeling them out of the survey in case their quota is already full.

Supplemental Material >

2.4. Managing the Survey

When administering your survey, you need to carefully monitor the entire process to avoid issues you may not have noticed during the design phase. Some suggestions are given below.

- Soft-launch the survey. Before launching the full-scale survey, you should run a small-scale version, or "soft launch," of the complete survey. This is slightly different from the pretesting and piloting discussed in **Supplemental Appendix A-3**, which is about testing the content and questions. It is about figuring out whether there are technical issues with your survey flow.
- Monitor the survey. One advantage of online surveys is that you can monitor the data collection in real time and adapt to unforeseen circumstances. First, you must pay attention to dropout rates. If you notice respondents dropping out at particular points, you may want to pause the survey and figure out the problem. This will also help you flag potential technical issues you may have missed while testing. Similarly, monitor your quotas. If one quota is filling up too fast, it will be challenging to fill the other groups later on. Finally, regularly check the designated survey email inbox in case respondents have sent emails that flag problems.
- Check the data during the collection process. From the earliest responses, you should have a procedure to start checking the validity of answers, tabulating answers, and spotting possible misunderstandings or errors. Also, check that the data you are collecting are being recorded correctly.

2.5. Attrition

This section provides some advice on dealing with attrition.

2.5.1. Reporting attrition. The level of attrition and its correlation with observable and unobservable characteristics are important issues in a survey. It is good practice to report detailed statistics on attrition for your survey, including (*a*) your total attrition rate with a clear definition (e.g., which respondents count as "having started the survey" versus "completed" it? Do you count respondents who failed possible attention checks? Or who skipped the basic demographic questions?); (*b*) your attrition rate at key stages in the survey, such as upon or after learning the topic of the survey, answering socioeconomic questions, seeing an experimental treatment, etc.; and (*c*) correlations of attrition with respondent characteristics. To be able to test for differential attrition, some background information on the respondent is needed. If there is no outside source for that information (e.g., administrative data), there is a strong rationale for asking socioeconomic and background questions earlier in the survey to see whether respondents are selectively dropping out. There are trade-offs in this ordering of survey blocks, which are discussed further in **Supplemental Appendix A-3.6**.

Supplemental Material >

Supplemental Table A-3 gives a sense of the distribution of attrition rates across various papers and platforms. Subject to the caveat that attrition is not defined in the same way across different studies, attrition rates tend to range between 15% and somewhat above 30%, depending on the platform used and the survey length. Patterns of correlation between personal attributes and attrition are not clear-cut and will likely depend on the topic and design of the survey. In a study across 20 countries, Dechezleprêtre et al. (2022) find that female, younger, lower-income, and less-educated respondents are more likely to drop out, but differences in attrition rates are not large. **Supplemental Appendix A-1.4** shows that survey length may be correlated with higher attrition. Respondents in the treatment branches of surveys with an experimental component are sometimes more likely to drop out, either because of the added time commitment or based on the topic (which can introduce bias in the treatment effects estimated).

2.5.2. Preventing attrition. The best remedies for attrition are a smooth respondent experience (e.g., pages loading quickly, a clear visual design, and well-formulated questions, as described in **Supplemental Appendix A-3**), a shorter survey, and good incentives (here again, it helps if survey companies have a variety of possible rewards that appeal to a broad range of people, rather than just one type of reward, which may induce selection). It is a good idea to avoid too many attention check questions (see **Supplemental Appendix A-2**), personal questions, and complex questions, all of which could irritate respondents. It is also good to be careful about revealing the topic too early on before you know enough about who the respondents are, so that you can check for selective/differential attrition based on the topic.

2.5.3. Correcting for nonresponse bias (selection and attrition). There are different types of methods to correct for nonresponse bias, which are described in detail in **Supplemental Appendix A-1.5**. These include reweighting responses (Section A-1.5.1), modeling selection directly (Section A-1.5.2), bounding the effects of interest (Section A-1.5.3), and imputing missing data (Section A-1.5.4).

2.5.4. Best practice tips. The first step in dealing with selection and attrition is accurately reporting them to your readers. For attrition, (*a*) describe your overall rate of attrition, (*b*) correlate it with observables, and (*c*) provide the timeline of when people drop out (see Section 2.5). For selection, compare your sample carefully to the target population along as many dimensions as possible. If the characteristics are similar, this is reassuring, although responders may differ from nonresponders in other ways that are not measurable (and this is not testable). For item nonresponse, you can identify specific questions where there are more or many missing responses. For example, if you have to use a variable with many missing observations, you need to discuss this more extensively than if the variable has only a few missing responses. Your adjustments or corrective procedure and reporting should depend on the magnitude of the nonresponse, selection, and attrition problems. It may be worthwhile checking the robustness of your results to various correction methods among those described in **Supplemental Appendix A-1.5**.

It is helpful to report your raw survey results before any adjustment, either as a benchmark case or in an appendix. You can acknowledge that the results hold, strictly speaking, just for your sample and may or may not hold for the target population. After you apply one or several correction methods (reweighting, bounding, imputation, or model-based adjustments), you can report these results (in the main text or an appendix) for comparison with the raw ones. A final tip is to use questions on attitudes, views, or beliefs from existing, high-quality, representative (of your target population) surveys that can serve as benchmarks. You can compare the answers in your study to those in benchmark surveys so that you have an extra validation beyond comparing socioeconomic or demographic characteristics.

3. RESPONSE BIASES

This section reviews methods for detecting, minimizing, and dealing with different response biases that can arise in surveys.[10] Sources of biases include the respondents' behavior (e.g., carelessness or SDB), the content of the question (e.g., leading questions), the design of the questionnaire (e.g., the order of questions that can induce priming), and the characteristics of the survey situation itself (e.g., EDE). The section covers biases related to the choice of answer options that are unrelated to their content (moderacy bias, extreme response bias, and ordering bias), acquiescence bias, EDE, and SDB.[11] The first line of defense against these biases is, once again, proper survey design. Good design avoids inducing biases (e.g., by using neutral rather than slanted questions) and reduces survey fatigue and annoyance, which can exacerbate the likelihood of all these biases occurring. Accounting for these biases may be particularly important in cross-country studies and when comparing different groups within a country, since the tendency to respond in a given way may vary across cultures and within a country by socioeconomic and other factors.

3.1. Biases in Answer Selection: Moderacy, Extreme Response, and Response Order Biases

There are three biases related to systematically picking a given type of answer option regardless of the content of the question (Bogner & Landrock 2016). They may occur out of satisficing or carelessness. Krosnick et al. (1996) suggest that these biases occur as respondents try to take shortcuts to minimize the cognitive load. They may also be natural consequences of how we process information based on the serial position of alternatives and their visual presentation. These biases are the following.

- Moderacy response bias is the tendency to respond to each question by choosing a category in the middle of the scale.
- Extreme response bias is the tendency to respond with extreme values on the rating scale.
- Response order bias occurs when the order of response options in a list or a rating scale influences the response chosen. The primacy effect occurs when respondents are more likely to

[10]This section is not about psychological biases in general (e.g., lack of understanding of probabilities, overestimation of rare events, other fallacies, etc.). It is about biases precisely due to the survey setting or question design.

[11]This is not an exhaustive list of biases, but it does cover the most important ones. Other biases include hostility bias (responses due to feelings of anger from the respondent, e.g., if forced to complete a long survey or the questions are upsetting) and sponsorship bias (whereby a respondent is influenced by the perceptions of the person or organization conducting the survey), which are addressed by similar methods to the ones related to SDB (see Section 3.3) and EDE (see Section 3.5), as well as by the general advice on question design (see **Supplemental Appendix A-3**) and recruitment (see Section 2.3).

Supplemental Material >

select one of the first alternatives provided, and it is more common in written surveys. This tendency can be due to satisficing, whereby a respondent uses the first acceptable response alternative without paying particular attention to the other options. The recency effect occurs when respondents choose one of the last items presented to them (more common in face-to-face or orally presented surveys).

Detecting these biases is not easy. A given answer pattern can arise because of carelessness combined with a particular reaction (e.g., picking the middle option versus picking the first available option) or because a respondent may legitimately have extreme or uninformed/neutral views on a given topic. Incidentally, these biases can be difficult to disentangle from each other. For instance, systematically picking the first answer option (order bias) may look like extreme response bias, depending on the content of these answer options and their ordering (e.g., the first option may systematically be an extreme one), especially in the case of ordinal closed-ended questions. Nominal closed-ended questions are most likely to be prone to order bias.

Supplemental Material >

3.1.1. Possible solutions. Solutions for detecting and correcting these biases come in three shapes: design solutions, reduced-form solutions, and model-based solutions. Design-based solutions, described in the best practices in **Supplemental Appendix A-3**, involve keeping the cognitive burden to a minimum to reduce the risk of satisficing. Customized scales and answer options scales with differentiated options are essential. For instance, three-point answer scales will almost mechanically lead to situations that look like extreme response bias or moderacy bias because there are too few differentiated options. You should not eliminate a truly neutral middle option for fear of moderacy bias. Instead, informative options should reduce the likelihood that respondents pick middle answers due to a lack of alternatives.

Reduced-form solutions involve constructing an index measuring the extent of each problem for a given respondent. For extreme response bias, an easy method is to build an extreme response sum-score (ERS) index (Johnson et al. 2005). For each variable potentially subject to the bias, you can create a dichotomized variable equal to one if the answer is an extreme value and equal to zero otherwise and then sum these variables. The ERS index could potentially serve as a control in the analysis. The validity of such reduced-form approaches is greater if the underlying items are not too similar in content. Otherwise, such measures mix substantive issues with bias, as it could well be that the respondent has extreme views on a set of closely related topics. Similar approaches can be taken for moderacy bias. For each of these biases, there can also be specific sets of questions designed to explicitly measure them (e.g., ask about unrelated issues and see whether respondents select similar options despite the differences in content). Still, such questions increase the survey's burden and can appear weird to respondents.

Model-based solutions consider the response style, substantive factors, and content of each question. Examples of these models are listed below.

- De Jong et al. (2008) use an item response theory (IRT) model that assumes a continuous and stable extreme response style while allowing items to be differentially useful to measure it. The key drawback is that it uses items in their dichotomized form.
- Byrne (1989), Byrne et al. (1989), and Jöreskog (2005) use a multi-group confirmatory factor analysis (CFA) framework to assess whether different response styles affect group comparisons.
- Cheung & Rensvold (2000) develop a multi-group structural equation model (SEM) that allows them to check whether group comparisons are invalidated by response style, but it does not directly detect or correct for it.

- Billiet & McClendon (2000) and Welkenhuysen-Gybels et al. (2003) develop an SEM specification to correct for acquiescence bias, but this is not applicable to extreme response styles as the relationship of the response outcome with the response style is nonmonotone.
- Morren et al. (2011) use a latent class factor analysis (LCFA) model whereby the response to each item is explained through a multinomial logistic regression on substantive factors and the individual ERS index. This specification, where the coefficient attached to the ERS variable represents the likelihood of an extreme answer, is used to deal with the U-shaped relationship of ERS with response outcomes (i.e., ERS can imply either a very high or a very low answer).

3.1.2. Specific solutions for response order bias. There is a range of solutions specifically for response order bias.

- Avoid long response lists. Response order bias is more likely to occur when respondents need to read through long lists of alternatives.
- Use seemingly open-ended questions, as suggested by Pasek & Krosnick (2010). These questions separate the question stem from the response alternatives with a short semantic pause to encourage individuals to stop and think before answering the question, almost as if they were answering an open-ended question. For instance, instead of asking, "If more revenues were needed to finance transfers to low-income households, would you prefer that the personal income tax or the corporate income tax was increased?", response order effects can be reduced by asking, "If more revenues were needed to finance transfers to low-income households, which tax would you rather increase? [pause] Would you rather increase the personal or the corporate income tax?".
- Randomize the order of response options for questions with unordered (nominal) response options or invert the order for ordinal questions. This is mainly to detect rather than to solve for response order bias and could deceive even nonbiased respondents who try to answer survey questions efficiently if they are caught by surprise by a different order.
- Prefer forced-choice response formats to check all response formats (see **Supplemental Appendix A-3**)—i.e., essentially, ask respondents to evaluate each response option rather than a list of alternatives.

Supplemental Material >

3.2. Unintended Question Order Effects

I now discuss unintented question order effects.

3.2.1. Why do question order effects occur? To deal with unintended question order effects, it is useful to consider why they arise in the first place. We can distinguish between cognitive-based order effects and normative-based order effects (Dillman et al. 2014). Cognitive-based order effects include (*a*) priming, that is, bringing to mind content that becomes more salient in subsequent questions; (*b*) carryover, that is, answering two questions that appear related using similar criteria and thought processes;[12] and (*c*) anchoring, that is, the standard applied to one question shapes the standard applied to a second one.

Normative-based order effects include the wish to appear (*a*) evenhanded or fair, so that respondents will adjust their answers when asked about two different situations, groups, policies,

[12]An example from Dillman et al. (2014) is that when people were asked "How would you describe your marriage?" and "How would you say things are these days?" answers to these two questions varied greatly depending on which was asked first because of carryover effects.

etc. (and answer differently than if they had been asked about only one of them or in a different order); (*b*) consistent, whereby respondents prefer that answers to a later question be consistent with answers to a previous one; and (*c*) moderate, whereby respondents who try not to appear extreme will tend to reject some items and support others. The literature distinguishes between assimilation, or making answers across questions more similar than they would otherwise be, and contrasting, or making answers more different. In general, it is difficult to pinpoint the direction of the bias induced without further case-specific information.

3.2.2. Possible solutions. First of all, it is important to consider carefully the order of questions when designing the survey. There is no general solution, but you must be aware of these effects. If you worry about order effects and it is possible to vary the order of questions without destroying the logical flow of your survey, it makes sense to randomize the order of individual questions or question blocks. However, be aware that later questions suffer more from survey fatigue, so try not to conflate order randomization by shifting the questions or blocks too far across the survey. Suppose you worry about order effects between two questions and want to prevent respondents from associating them. In that case, you should try to visually dissociate them, for example, by placing them on different survey pages or spreading them out in the questionnaire. The design of the survey with order effects in mind is bound to be an iterative process: Once you have focused on arranging questions to reduce order effects, make sure you have not disturbed the logic and consistency of the questionnaire.

3.3. Social Desirability Bias and Solutions

SDB typically stems from the desire of respondents to avoid embarrassment and project a favorable image to others, resulting in respondents not revealing their actual attitudes. The prevalence of this bias will depend on the topic, questions, respondent, mode of the survey, and the social context. For instance, in some circles, anti-immigrant views are not tolerated, and those who hold them may try to hide them. In other settings, people express such views more freely.

Overall, there are some general issues to take into account in online surveys. The setting of online surveys likely minimizes SDB because there is no surveyor in front of the respondent or on the phone. The social context equivalent in an online survey relates to (*a*) the revealed identity of the surveyor or entity running the survey, (*b*) the level of anonymity provided to the respondent, and (*c*) the knowledge of what questions will be used for. These three aspects can be controlled and tested for. Regarding the identity of the surveyor, it is important to be very aware of what you reveal to the respondent (e.g., "nonpartisan researchers" versus "researchers from the Economics Department at Harvard") and, more generally, to consider the issues raised about recruiting respondents in Section 2.3. Both (*a*) and (*c*) can be tested thanks to questions at the end of the survey, asking about the attitude toward the surveyor or entity (e.g., Do respondents think they are biased in a particular way? Do they have favorable or unfavorable views of the surveyor?) and questions about the perceived purpose and intent of the survey.

Respondents should be assured of complete anonymity in the survey landing and consent page (see Section 2.3). However, if your survey contains sensitive questions, it can be helpful to reemphasize that their answers are confidential and anonymous before asking a particular set of questions. You can strategically place sentences such as "As a reminder, all of your answers on this survey are confidential and anonymous" before sensitive questions. However, do not overdo this as it has diminishing returns if used more than once or twice in a survey. More generally, you will need to be careful not to have too many sensitive items, as no method can prevent SDB in that case. These items must be placed strategically in your survey (see **Supplemental Appendix A-3.6**).

Supplemental Material >

The methods to deal with SDB reviewed in **Supplemental Appendix A-4.1** have some disadvantages. Some may only help if SDB arises out of the wish to appear in a certain way to the surveyor or others and may not be effective in reducing SDB that is due to self-image concerns. As a general rule, the less directly respondents have to answer a problematic question, the more you can minimize the self-image bias. Furthermore, some questions may be too sensitive, and no method may work. Some of these methods can only identify group-level distributions in the answers to the sensitive questions but not respondent-level responses (only probabilistically). Some methods involve questions that are hard to understand and involve lengthier instructions, which respondents may find strange and confusing. For some questions, respondents need to understand the procedures and underlying logic to understand they are protected and trust the anonymity. Regardless, even with these methods, some may be reluctant to answer sensitive questions. Respondents may still worry that surveyors can infer their views (particularly in the randomized response technique or crosswise technique). **Supplemental Appendix A-4.1** describes several methods to dampen SDB and provides examples.

Supplemental Material >

3.4. Acquiescence Bias

Acquiescence is the tendency to answer items in a positive way regardless of their content, for instance, systematically selecting categories such as "agree," "true," or "yes" (Billiet & Davidov 2008).

There is evidence that acquiescence bias correlates with difficult questions, survey fatigue, and lack of knowledge or interest in a topic (Krosnick 1999). Answering in a positive manner may be a way out at a low cognitive cost for the respondent and requires less effort than having to think of the pros and cons of an issue. It is debatable whether it represents an individual trait or is rather due to the question or survey itself.[13]

Some of the ways to prevent acquiescence bias are already covered in the question design in **Supplemental Appendix A-3**. Some examples are given below.

- Do not ask ambiguous, unclear, or complicated questions. Respondents may be tempted to agree as an easy way out instead of trying to guess what you mean.
- Ultimately, the most robust way to decrease the likelihood of acquiescence is to avoid questions of the type agree-disagree, true-false, and yes-no (Krosnick 1999). Instead of agree-disagree questions, ask questions with direct, item-specific scales, in which the categories used to express the opinion are precisely adapted to the item. For instance, instead of asking "Do you agree or disagree that your health is good?" with answer options "agree completely, agree somewhat, neither agree nor disagree, disagree somewhat, disagree completely," change the question to "How would you rate your health?" with answer options "very bad, bad, neither good nor bad, good, very good."
- Instead of yes-no and true-false questions, change the type of question to offer answer options that include all possible views: For example, instead of asking "Do you think that student loans should be forgiven for everyone?" with answer options "yes, no," you could ask "Do you believe that student loans should or should not be forgiven for everyone?" with answer options "they should be forgiven, they should not be forgiven."

[13] This bias is likely more prevalent in face-to-face or phone surveys, where the surveyor may not be able to keep an entirely neutral tone, than in online surveys. Note that, unlike the EDE below, it may not matter whether the respondent knows the purpose of the study, since acquiescence bias reflects a tendency to agree with statements rather than to try to please the experimenter.

- In the case of bipolar questions, you should always use a balanced scale that includes equal numbers of positive and negative options.
- You can consider having two versions of fundamental questions in your survey, one with a positive statement and one with a negative one, to test for bias. This only works if positive and negative items in a scale are equally likely to be subject to acquiescence. It also requires burdensome negative or double negative formulations, so you should use them sparingly and not one after the other. A more general version of this advice is that if you ask for people's opinions on some topic or issue, you could try to have half the items phrased in a "pro" direction and the other half in a "con" direction.
- At a group level (e.g., within a country, within an age bracket, etc.), you can randomize who sees a given question and who sees its inverse to get an estimate of acquiescence bias (see Dechezleprêtre et al. 2022).
- The reduced-form and model-based approaches outlined for extreme response bias in Section 3.1 can be adapted to acquiescence bias as well (with their advantages and shortcomings).

3.5. Experimenter Demand Effect

EDE refers to the fact that respondents who are in the treatment branch may differentially form views about the experimenter's expectations compared to those in the control group. Thus, their responses may reflect other considerations in addition to the actual treatment effect. I would broaden that definition to "surveyor demand effect," which means that, more generally, respondents may feel that the surveyor wants them to respond in a given way to some questions of the survey. In this case, the threat is to the validity of descriptive statistics too (not only treatment effects). This section closely follows the work of Haaland et al. (2020) and Stantcheva (2021). Because EDE shares similarities with SDB, it is useful to think of the general advice (related to anonymity, the identity of the surveyor, and the purpose of the survey) discussed in Section 3.3.

Possible solutions to EDE are listed below.

- Anonymity is one way to minimize EDE. Online surveys can relieve social pressure since respondents are anonymous and take the survey on their computers or phones with complete privacy. Some of the solutions to EDE can be found in the section on SDB, since SDB partially stems from thinking about what is socially desirable to the surveyor.
- Monetary incentives and real-stakes questions could help motivate respondents to answer accurately. As discussed in **Supplemental Appendix A-3.7**, monetary incentives for truthful revelations can only be used for questions that have a correct answer, while real-stakes questions can be used for attitudes and policy views.
- Obfuscated follow-ups can be used to estimate treatment effects, as suggested by Haaland & Roth (2020, 2023). These are follow-up studies with the same respondents as in the initial experiment, where dependent variables are elicited to estimate treatment effects, but without the respondents knowing that the original and follow-up surveys are related.
- Obfuscated information treatments try to obscure the purpose of the experiment. Three ways of doing this are (*a*) to provide respondents with additional pieces of information that are irrelevant to the actual goal, (*b*) to ask respondents questions or give them tasks on unrelated issues, and (*c*) to give people an unrelated (but valid) reason to explain why they receive the information of interest.
- Design and question wording are important, in particular the advice on neutral and balanced framing from **Supplemental Appendix A-3**.
- Hiding the purpose of the experiment or the study from respondents can help alleviate EDE. Having different blocks and focusing on different angles in the survey can help in that regard.

Supplemental Material >

For instance, in the survey by Stantcheva (2021), respondents are walked through multiple different blocks that have different foci: open-ended questions about their main concerns and shortcomings of the tax system, factual knowledge questions, questions on the efficiency costs or distributional impacts of taxes, and, only at the end, questions about their preferred policies. On balance, because different aspects are touched upon, the questions do not appear to lean either in favor of or against taxes.

- Similarly, measuring beliefs about the study purpose should be done at the end to avoid impacting subsequent answers. It is, in general, good practice to have feedback questions and entry boxes at the end of the survey to assess whether respondents thought the study was biased and, if yes, in which direction.
- Demand treatments explicitly introduce (possibly randomized) questions that use explicit signals of the surveyor's wishes and use them to measure the extent of EDE. In particular, one can compare the answers to self-reported (more likely to be sensitive to EDE) versus real-stakes (less likely to be sensitive to EDE) questions. Allcott & Taubinsky (2015) argue that if demand effects are driving behavior in experiments, then they should be more pronounced for respondents who are more able to detect the intent of the study and are more willing to change their choices given the experimenter's intent.

4. SURVEY EXPERIMENTS

A survey experiment is simply an experiment embedded in a survey. You do not necessarily need an experimental part in your survey. The outcomes of interest are the endogenous variables; the experimentally manipulated conditions are the independent variables. Respondents can be part of a control group or one of several treatment groups in which they are subject to some experimental treatment.

There are multiple benefits to survey experiments. Similar to other experiments, they allow you, in principle, to test for and identify causal relationships. They are very flexible in their design and process and come in many different forms, described below. Because you have control over creating the questionnaire, you can measure precisely the input and dependent variables needed. Finally, you can sample the type of respondents most suited for your research question (e.g., sampling for range, representation, expertise, etc.) and include a wide variety of people in your experiment.

4.1. General Considerations

Some general considerations related to survey experiments are as follows.

4.1.1. Challenges in survey experiments. Survey experiments face some of the same challenges as other types of experiments (e.g., lab or field experiments) and some specific ones. First, there is a risk of confounding (by unobservable factors) and pretreatment contamination (i.e., respondents may experience a similar treatment outside of the experiment, which means that the treatment and control groups will not show marked differences in outcomes, even if the underlying causal effect of the treatment on the outcome exists).

The problem of information equivalence is more specific to survey experiments. Different respondents can interpret an experimental intervention designed to shift beliefs or information sets differently, resulting in them de facto experiencing a different treatment (Dafoe et al. 2018). In other words, an experimental treatment meant to trigger a given construct A could also trigger an unintended construct B. Suppose B is not experimentally manipulated or held constant at the level of analysis in the experimental design. In that case, we cannot disentangle the effects of A and B on the outcome. In the case of survey experiments, a treatment might unintentionally create associations with correlated elements from the respondents' real lives, preventing us from estimating the effect of the manipulated factor separately from that of the confounder. An assumption required to estimate the causal impact of a given factor of interest is that the survey manipulation be information equivalent with respect to the background features of the scenario. However, manipulating respondents' beliefs about a given factor will often affect their beliefs about other background factors in the scenario, too. An example given by Dafoe et al. (2018) is that labeling a country a democracy affects respondents' beliefs about where the country is located and about its demographic characteristics. Information equivalence is similar to the exclusion restriction in instrumental variable (IV) designs. As a result, it is generally difficult to use a survey treatment as an IV for an endogenous variable.[14] More often than not, we focus on the reduced form of the treatment on outcome variables.

An additional concern in survey experiments is whether the treatment is supposed to mimic a real-world treatment, e.g., the information we could see in the media, or whether it is more abstract. External validity is also a concern based on the sample composition and setting (see Section 2).

4.1.2. Different types of survey experiments. There are many different types of survey experiments. In brief, information treatments work by correcting or expanding respondents' information via learning and updating. Priming treatments activate certain mental concepts or mindsets or make certain features more salient than others. Vignette designs and factorial experiments modify various attributes of the choice context in a controlled manner to study their impacts on judgment and behavior.

4.1.3. Design choices. Regardless of your treatment type, you need to make some decisions on design, particularly on two interconnected issues: Will you use a between- or within-respondent design? When will you measure your dependent variable of interest: before, after, or both before and after the treatment?

Between-subject designs are those in which each respondent is only subject to one experimental condition. For example, a treatment group sees one video, and the control group sees no video. In within-subject designs, each respondent is subject to multiple experimental conditions (not necessarily all). The only difference between the groups is the order in which the conditions are administered, which allows researchers to rule out confounds between time and the treatment. For instance, both groups see the video, but at different points in the experiment.

There can be multiple dependent variables of interest. In information or pedagogical experiments (and sometimes in other types of experiments too), there is an extra consideration: There are dependent variables that can be labeled first-stage variables—i.e., the belief, information, or knowledge that your treatment is trying to shift (e.g., the perceived share of immigrants in the respondent's country)—and second-stage variables, which are dependent variables influenced by those first-stage ones (e.g., policy views such as whether there should be more immigration). Another example comes from Stantcheva (2021). The perceived mechanisms of income taxes (their efficiency effects or distributional impacts) are first-stage variables. In contrast, the preferred level of the top income tax rate is a second-stage variable.

Table 1 illustrates different experimental designs. Each group of two rows is one experiment, and each row represents a randomization branch within the experiment, denoted by groups 1 and 2. "Treatment" refers to when the respondent is exposed to treatment. Measurements of the dependent variables are represented by the letter O; Q denotes measurements of closely related

[14]This is not that different from field randomized controlled trials or other types of treatment.

Table 1 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	O_1	Treatment	O_2
	2	O_3		O_4

O = observation of the dependent variables (first and/or second stage).

Q = observation of variables closely related to the dependent variables (first and/or second stage).

T_1 = point in the survey where the first measure is possibly taken.

T_2 = point in the survey where the second measure is possibly taken.

but not identical variables. Note that the measurements O and Q can represent measurements of the first- or second-stage variables or both.

The first three experimental designs represent between-respondent designs. In the posttest design, which is very standard, one group is exposed to treatment, the other one is not, and the dependent variables are measured after the treatment only. The dependent variables are measured twice, before and after the experiment, in the prepost design. In the quasi-prepost design, the dependent variables are again measured before and after, using similar rather than identical elicitations. Furthermore, one can imagine variations on these designs for information and pedagogical experiments, where, e.g., first-stage variables are measured before and after (prepost or quasi-prepost) while second-stage variables are measured only after the treatment (posttest). The within-subject design exposes the two groups to the same treatment but at different points in time and measures the dependent variables twice (and at the same time) for both groups. A within-subject design makes the most sense with prior measurement of the outcome variable for at least one group (e.g., measurement O_3 for group 2).

In general, eliciting posterior first-stage variables (after the treatment) is advisable. It is the only way to get an actual first stage of your treatment and see if it worked and, if yes, by how much. Furthermore, it is often of interest per se to estimate the effects of information on these first-stage variables (e.g., on knowledge), the direction and speed of learning, and the updating that people do.

There are also benefits of eliciting prior first-stage variables (before the treatment). First, this allows you to estimate heterogeneous treatment effects based on the prior value of the dependent variable. In particular, you will be able to check whether treatment effects are larger for those who, given their baseline values of the dependent variable, de facto received a larger shock from the treatment. It can also increase statistical power in conjunction with postmeasures (this is true for the measurement of second-stage variables too).

Dependent second-stage variables always have to be measured (at least) posttreatment. The question is whether you need to measure them pretreatment too. Pretreatment measures can add precision since they allow you to filter out an individual fixed effect and the tendency of a respondent to respond in a given way. As a first pass, a substitute is, of course, to control for covariates (measured pretreatment only ideally) that are relevant.

Yet, for both first- and second-stage variables, there are concerns about asking twice: (*a*) Consistency pressures may prompt the respondents to answer similarly pre- and posttreatment. Yet, it is possible to ask similar, but not identical, questions to avoid making the inconsistency salient and to vary the question format. (*b*) Asking twice about the same thing may also lead to more EDE or SDB if it makes the respondent more aware of the topic of the experiment. (*c*) Related to this, the biggest worry is perhaps that, by asking these questions, you are already priming respondents to think about the topic, and priming in itself may not be neutral. (*d*) More elicitations also mean more time commitment for the respondents, increasing survey fatigue and cognitive load and potentially leading to the problems discussed in **Supplemental Appendix A-2**. The trade-off will generally depend on the topic and on how sensitive and prone to priming, EDE, and SDB it is. If performing a prior measurement, it may be advisable to use similar but not identical questions.

Supplemental Material >

Some of the benefits and shortcomings of the within-subject design—since it often goes hand-in-hand with measuring dependent variables before and after the treatment for at least some respondents—are the same as for the elicitations just discussed (e.g., they increase statistical power). In addition, a within-subject design lets you control for time and duration effects or order effects in the survey itself. In **Table 1**, if the order of conditions were not randomly assigned, there would not be two distinct experimental groups, and any effect of time or repeated measurement of the dependent variable would be confounded with the treatment. Because respondents receive multiple treatments, they may require smaller samples to achieve sufficient power.

For any of these designs, a good elicitation of first- and second-stage variables goes through proper question design, as explained in **Supplemental Appendix A-3**.

4.2. Priming Treatments

Priming treatments activate mental concepts or mindsets through subtle situational cues. Typical priming techniques include actively prompting subjects to think about specific concepts or, more subtly, subjecting them to visual or other stimuli. Researchers can use priming to measure implicit attitudes without the respondent's knowing what is being measured. This lack of awareness makes measured effects less likely to be distorted by the biases described in Section 3. In the context of priming experiments, the goal is to trigger the relative salience of a concept and mindset to measure its causal effect on outcome variables of interest. Priming treatments can extract relationships between different constructs even when the respondent may not be aware these relationships exist. Cohn & Maréchal (2016) review the literature on priming in economics, and Bargh & Chartrand (2014) offer a theoretical framework and review the literature in psychology.

4.2.1. Mechanism and theoretical aspects. Priming can change the relative weight (or salience) individuals attach to the primed concept at a given moment (Cohn & Maréchal 2016). Hence, priming acts differently from more explicit treatments such as information treatments, since it exploits a subconscious exogenous increase in salience with respect to the primed element. Often, priming is concerned with unintended effects of environmental forces on feelings, behaviors, and attitudes, which individuals may not be aware of. Following Bargh & Chartrand (2014), experiments can focus on (*a*) conceptual priming, which activates mental concepts in one context so as to induce mental representations that are used subconsciously in subsequent contexts; or on (*b*) mindset priming, which primes a given way of or procedure for thinking (i.e., a mindset) by having the participant actively use that procedure. For instance, Cohn et al. (2015) prime financial professionals to think in a risk-averse way (a risk-averse mindset), while Alesina et al. (2023) prime respondents to think about immigration (a concept). Some experiments prime identity by highlighting specific groups the respondent is part of.

Two channels through which priming works are accessibility—i.e., by making some features more salient, knowledge stored in memory can be reactivated for a judgment task—and applicability, that is, the degree to which the presented stimulus or stored knowledge is perceived as applicable to another context (Althaus & Kim 2006). Scheufele (2000) describes the theoretical difference between priming and framing, arguing that the former works through the channel of accessibility, while the latter works through prospect theory (i.e., leading to a different cognitive scheme to interpret the issue). For instance, framing may change the criteria according to which a policy is judged.

An important distinction to keep in mind when priming is that certain dimensions may be context dependent, while others are hardwired in how we think. This can explain why certain priming interventions work (as they act on the dimensions that are context dependent) and others do not (as they act on those that are hardwired). In other words, certain dimensions of our thinking, identity, or mindset are already salient before the priming treatment. If the treatment tries to emphasize them, it may be difficult to detect an additional effect. If the treatment tries to dampen them, it may not produce meaningful change.

4.2.2. Types of priming. Priming can rely on different textual, audio, or visual modes. Some types of priming techniques used in the literature include the ones listed below.

- Slanted questions. Kuziemko et al. (2015) prime respondents to think negatively about the government by asking them questions about issues they dislike, such as the Citizens United Campaign or the Wall Street bailout. Stantcheva (2022) primes respondents to think about their benefits and costs from international trade as consumers versus employees by asking them a series of questions about how trade has impacted their consumption and labor market experience.
- Order randomization/changing order of questions. This is done, for instance, by Alesina et al. (2023), who randomize the order in which respondents are asked questions about immigrants versus questions about redistribution to test the effect of immigration perceptions on views on redistribution.
- Use of words or names with different connotations. One method of priming is presenting respondents with hypothetical scenarios in which the names of the people are varied. Names are chosen so as to evoke specific ethnicities or nationalities [e.g., Alesina et al. (2023) randomize the name of immigrants used in hypothetical examples]. One can also use words with different connotations, without the respondent's being aware of this. For instance, Merolla et al. (2013) test whether support for the DREAM Act and birthright citizenship changes when the questions refer to immigrants as "unauthorized," "undocumented," or "illegal."
- Varying the illustrations and images shown alongside the information. Another priming method involves showing respondents the same information but with different illustrations. For instance, Kuziemko et al. (2015) provide respondents with information about the (low) share of households who pay the estate tax. However, respondents who receive this information are split into two treatment groups. One group has a picture of a mansion on the page, alongside the information, while the other does not (see **Supplemental Figure A-16**). Thus, respondents who see the mansion picture are primed to think about the lifestyles of the wealthy while also receiving the information, and the differential treatment effect between these two groups can be attributed to that prime. Brader et al. (2008) prime the racial identity of immigrants by presenting respondents with a pseudoarticle from the *New York Times* about an immigrant, varying the appearance of the immigrant depicted in the picture that accompanies the article.
- Priming through images. Specific images can be used to prime respondents. For instance, Israel et al. (2014) prompt respondents to think about vacation or old age using images and text. Their goal is to influence respondents' time preferences and discount rates. Vacation scenes tilt time preferences toward the present, while pictures of old age increase the discount factor.

Supplemental Material >

- Priming through videos. Videos can prime respondents as well, sometimes in an immersive way. For example, Guiso et al. (2018) make respondents watch a horror movie to test whether fear increases risk aversion.

Note that, with some of these techniques, it may be difficult to disentangle the effect of priming from that of information provision. A prime should ideally not lead to any learning or belief updating, since it would then be impossible to disentangle the effect of the prime from the effect of the new information. It is also possible to probe whether respondents are aware of the prime at the end of the survey, by asking them about their perceived purpose of the survey and whether they noticed links between different parts. **Supplemental Appendix A-5.1** provides examples of papers using different priming techniques in survey experiments. There is a bit of recent controversy on the soundness of priming studies in the cognitive psychology literature (see **Supplemental Appendix A-5.1**).

Supplemental Material >

4.3. Information and Pedagogical Treatments

Information provision experiments exogenously change the information sets of respondents. Pedagogical treatments are closely related but go beyond providing information and facts and provide explanations and statements of how something works. For instance, Stantcheva (2021) explains to respondents how progressive taxes impact different people and what their efficiency costs are; Dechezleprêtre et al. (2022) provide explanations of how three critical policies to fight climate change can reduce emissions and benefit households with different income levels; Stantcheva (2022) provides treatments to respondents explaining the impacts of trade and trade policy on consumers and workers. Information or pedagogical treatments typically allow testing for the effects of specific information on outcomes, such as policy views or individual choices. They permit studying the impact of correcting misperceptions and checking belief updating.

4.3.1. Types of information and pedagogical treatments. There are many different types of information and pedagogical treatments, some of which are listed below. **Supplemental Appendix A-5.2** provides a review and examples of papers using these various methods.

- Quantitative information. Quantitative information can be precise and clear and minimize differences in interpretation across respondents. Yet, such treatments may be harder to understand and less appealing to respondents. For instance, Alesina et al. (2023) tell respondents in one treatment group the share of immigrants in their country and compare it to the share of immigrants in the countries with the highest and lowest immigrant shares in the OECD. Alesina et al. (2021) show respondents the evolution of the earnings gap between a Black man and a white man since the 1970s.
- Qualitative information. Treatments can provide more qualitative information, which is sometimes more effective than exact numbers and better suited to the question you are trying to answer. Qualitative information can also help make the treatment more homogeneous when doing an experiment in several countries or settings. For instance, Alesina et al. (2018) show respondents from five countries an animation on the lack of social mobility in their country. Without giving exact numbers, the treatment nevertheless gives the impression that few children born in the bottom of the distribution will move up in position.
- Anecdotes, stories, and narratives. Treatments can also take the form of anecdotes, stories, or narratives. For instance, Alesina et al. (2023) show respondents in one of the treatment groups an animation about a day in the life of a hardworking immigrant. Alesina et al. (2021) show a group of respondents a video about the differences in opportunities of a white child

and a Black child and use it to explain the deeper historical roots and consequences of systemic racism.

Pedagogical treatments may provide a mix of these types of content. For instance, an explanation may be bolstered by some concrete numbers and an example anecdote.

4.3.2. Form of the treatment. These treatments can be done through different media, including text, images, audio, videos, interactive exercises, and a combinations of these.

4.3.3. Additional dimensions of the treatment. There are some more variations in treatments to consider, listed below.

- Source of the information. Think about whether and how you want to inform respondents about the sources of the information. In some cases, this could increase the credibility of your information. The source per se may be part of the treatment and have its own effect. Indeed, there is evidence that the identity of the sender of information in experiments matters (see a review of papers in **Supplemental Appendix A-5.2**).
- Information specific to the respondent. You can also adapt the content of the treatments to be targeted to respondents, which can generate more attention and interest and be well suited for some questions. Kuziemko et al. (2015) show respondents where they would have been in the income distribution had inequality not increased since the 1980s. Hvidberg et al. (2021) show respondents where they rank in the income distributions of several reference groups, such as their neighbors or people with the same level of education.

Supplemental Material >

4.3.4. Methodological issues. There are some methodological issues to pay attention to when running information and pedagogical treatments, discussed below.

- The treatment should be relatively short. If it is a video or animation, do not force respondents to watch it before they are able to move on from the page. If you do this, some inattentive respondents may simply do something else while the video is running, and you will not be able to control for those who skip the treatment. Instead, encourage respondents to watch it and record the time spent on the page so that you can account for those who carelessly rushed through the treatment.
- Focus extensively on the design of your treatment, including the content and the format. Good graphics and appealing visual presentation are critical, especially given the quality of images, videos, or audio that people see every day on the web or social media. A treatment that appears poorly or nonprofessionally designed may generate negative reactions unrelated to the content.[15]
- The treatment should have a neutral tone (even if the information itself may not be neutral) and be easily understandable so as to avoid priming, EDE, and SDB.
- You have to be mindful of the trade-off between respondents' time and patience and the length and content of your treatment. For instance, videos take time to watch but can be vivid and powerful. Animations with text can be quicker to read and watch and can also be effective. Sometimes, simple text is enough to convey important information. For example, Kuziemko et al. (2015) simply tell respondents the share of people who pay the estate tax and show that this information significantly increases support for it.

[15] Projects done 10 years ago may have been very well designed for the standards of the time, but these standards change very fast!

- The more sophisticated your design, the harder it may become to avoid violating information equivalence and priming your respondents on some other dimension. For instance, if you want to convey information with videos that include people, the appearance and perceived identity of those people may not be neutral. The same goes for images or audio, as explained in the section on priming treatments (see Section 4.2).
- It is critical to elicit beliefs, perceptions, attitudes, and other first- and second-stage variables properly, which highlights the importance of the discussions around measurement in **Supplemental Appendix A-3.4**. Furthermore, you must adapt your belief elicitation to your treatment. For instance, asking only qualitative questions when the treatment conveys quantitative information may not be appropriate. Multiple measurements for your key variables are desirable.
- Make sure your experiment generates updating instead of priming. Ideally, you want to ensure that your treatment works through updating of beliefs and perceptions and not through mere priming (see Section 4.2). Common methods to mitigate concerns about priming include (*a*) measuring the first-stage variables (which your treatment manipulates) prior to the experiment in both the treatment and control groups, which ensures that both groups are similarly primed on the topic of interest; (*b*) introducing some separation between the experimental information or explanation provision and the elicitation of your dependent variables to ensure that short-term priming effects have dissipated; one possibility is to elicit your outcomes of interest in a follow-up survey rather than at the same time as the treatment, or, because recontacting respondents can be difficult (see Section 4.5), you can alternatively try to space out the experiment and the elicitation of outcomes in your survey; and (*c*) including an active control group, as described in Section 4.3.5 below.
- You can check whether respondents understood your treatment by using comprehension check questions. These questions should be adapted to address the key pieces of information in your treatment. For instance, there is no need to make them too difficult and ask about an exact number, if all that matters is that respondents got a general sense of the magnitudes. However, such questions are likely to signal to respondents the topic you are studying and the effects you may be interested in. Therefore, it may be a good idea to place them toward the end of your survey, or at least after eliciting your outcomes of interest.

Supplemental Material >

4.3.5. Active versus passive control groups. An active control group is a control group that receives different information on the topic of interest than the treatment group(s) [see, e.g., Bottan & Perez-Truglia (2022), who provide medical residents information on their ranking in the income distribution in different cities using two different data sources]. Because providing information takes time and attention and may generate emotions or thoughts that are not simply due to the information content, an active treatment group may offer a better comparison than a passive control group (that sees no information). Furthermore, since different respondents receive different information, there may be more variation caused by the treatment even for those with more accurate priors (i.e., the different treatments may shift the perceptions or beliefs of different people, with more or less accurate priors). The difficulty in using active control groups is that the information received is not neutral, so you cannot estimate the effect of the treatment group's information per se, which may be your goal. In addition, it may not be possible to find information that is different enough and yet truthful on the same variable of interest.

Related to this, you could have a mock treatment group that would see some unrelated information in order to make the total survey duration identical for the control and treatment groups (which is different from an active control group that receives different information on the same topic). It is essential, however, that the information received by the mock control group be truly neutral concerning the issue of interest, which can be difficult.

4.4. Factorial Experiments: Vignette and Conjoint Designs

Often also described as vignette experiments or conjoint designs, factorial experiments experimentally vary attributes and factors in hypothetical situations. A common design involves asking respondents to make normative judgments or hypothetical decisions in situations described in vignettes within which attributes and factors vary experimentally. The randomized variation of these attributes allows estimating their causal effect on responses.[16]

4.4.1. Types of factorial designs. Although the terminology is not clear-cut, factorial experiments can typically come in two formats. Vignettes are short descriptions or stories that vary across experimental conditions only along key factors of interest. They can be simple paragraphs of text describing people or situations but can also involve much more creative designs and media formats like images and videos. Conjoint designs often refer to tables or list descriptions of people and situations that only show attributes and their levels and avoid additional text. They are thus more direct, do not focus on storytelling, and can make specific features very salient. Both vignettes and conjoint designs can be in the form of simple designs (presenting a single profile or situation) or paired designs (presenting two profiles, which the respondent needs to rate or rank). **Supplemental Figure A-17** provides examples of single and paired vignette and conjoint designs.

Factorial designs can be compelling, as they allow you to examine the overall effects of a factor and its impact when presented in combination with other factors. Factorial designs are multiplicative (e.g., they can show combinations of attributes such as race and gender). They can also have high statistical power, as a small number of participants can evaluate many vignettes.

4.4.2. Benefits and challenges. One benefit of factorial designs is that they can present realistic, albeit hypothetical, scenarios. They can manipulate the effects of interest and present more complex scenarios while providing experimental variation. They allow us to test multiple hypotheses at once and to test for the effects of multiple treatment components separately, giving room to more complex behavioral explanations.

Another benefit is that factorial designs may limit the likelihood of SDB. When several (sometimes many) dimensions vary, it may be harder for respondents to know what is being sought after. In that sense, a conjoint design that lists characteristics may be more prone to SDB than vignettes, where characteristics can be smoothly hidden in stories.

One challenge of factorial designs is external validity. Would people make similar choices in real life as in hypothetical scenarios? Reassuringly, Hainmueller et al. (2015) compare the results from different conjoint and vignette designs with data from a referendum on giving foreign residents citizenship in Switzerland. They find that paired designs perform better in terms of predicting real-world voting. One possible explanation is that they lead to higher engagement, increase immersion, and reduce satisficing. A challenge is also the cognitive processes involved in these experiments, which may differ from those in everyday settings. Respondents simultaneously see different pieces of information that they may not otherwise see so clearly in everyday settings. This can also cause cognitive overload.

Furthermore, because the experiments are supposed to represent hypothetical but real-life scenarios, researchers must be careful to conceptually and theoretically specify all relevant

Supplemental Material >

[16]The use of a different type of vignettes, called anchoring vignettes, is discussed in **Supplemental Appendix A-3**.

dimensions to the situation and not omit any. A frequent criticism of vignette designs is that they miss important factors relevant to people's choices. At the same time, the number of cases in factorial designs can quickly become large since they are multiplicative.

4.4.3. Design issues and practical recommendations. The practical recommendations given below are useful for designing factorial experiments.

- Avoiding implausible combinations. When varying many attributes mechanically, some implausible combinations will arise but should be excluded (e.g., asking about a 30-year-old job seeker with 20 years of labor market experience).
- Keeping cognitive load manageable. Factorial designs can get tiring for respondents. Therefore, you need to keep the number of factors tested manageable and stick to a reasonable number of tasks (i.e., the number of vignettes or conjoint choices a respondent needs to make). There is no hard rule here, and it will depend on how long the rest of your survey and each vignette are.[17]
- Randomizing conditions. There are some decisions to make when designing the randomization of the vignettes or conjoint designs. Pure randomization is typically not ideal for experiments with many experimental conditions and where respondents are asked to make many choices. More often, there are ways to create different sets of conditions with better statistical and practical properties. For instance, one can create different sets of conditions to avoid having a respondent rate the same condition multiple times or only receiving randomizations along one dimension. "D-optimal" designs choose the sets of administered conditions that maximize statistical power.[18]
- Randomization of attribute order. Within each vignette or conjoint design, you can randomize the order of attributes to control for order effects. It is better to do this at the respondent level rather than at the question level to avoid cognitive overload (i.e., the respondent has to find information in different orders in each task).
- Advantages of table formats. Table formats have some advantages: They allow one to potentially randomize the order in which attributes are presented on a page in a way that natural running text cannot easily do. They may also be clearer and less tiring to read, especially if respondents are asked to perform many choice tasks. For a comparison between single and paired vignette and conjoint tables, readers are referred to Hainmueller et al. (2015).
- Within-subject design. In general, it is not advisable to have a pure between-subject design, where each person only sees one vignette or conjoint design. Instead, the within-subject design lets you control for respondent fixed effects.
- Mixed designs. In mixed designs, different groups of respondents will see different groups of vignettes. You can then make comparisons across respondents (since they see the same set) and also within respondents.
- Improve the level of immersion of respondents by using the appropriate media. Text is simple and may be the best choice in some settings. Images, animations, and videos can help in other settings.

[17] Bansak et al. (2018) find that response quality does not degrade up to 30 tasks in MTurk data.

[18] The reasons for using D-optimal designs instead of standard classical designs generally fall into two categories: Standard factorial designs require too many combinations for the amount of resources or time allowed for the experiment, and the design space is constrained (i.e., there are combinations of attributes that are not feasible or undesirable).

- Choose the right attributes. To avoid omitting essential variables, one can either focus on an attribute-driven approach, selecting features that are orthogonal to each other, or focusing on actual profile classes that are documented in the real world.
- Choose the setting. Particularly for vignettes, asking hypothetical questions about a setting close to a real-world environment can help participants feel more immersed and thus avoid satisficing.

4.4.4. Causal identification. Techniques to analyze the experimental results from vignettes include variance decomposition such as analysis of variance (ANOVA) or multilevel modeling (see Steiner et al. 2017). One commonly used quantity of interest in the conjoint analysis is the average marginal component effect (AMCE), which represents the causal effect of changing one profile attribute while averaging over the distribution of the remaining profile attributes (Hainmueller et al. 2014). For instance, a researcher may be interested in the AMCE of an immigrant's ethnicity that averages over the distribution of other immigrant characteristics such as age, education, or country of origin. Averaging over the distribution of other attributes can be more practical than conditioning on their specific values if many attributes are considered. However, the AMCE critically depends on the distribution used to average over profile attributes. **Supplemental Appendix A-5.3** reviews some methodological issues related to identification in factorial experiments.

Supplemental Material >

4.5. Follow-Up Surveys and Persistence

When doing survey experiments, you may worry that the treatment effects are temporary and will not persist. Treatment effects can dissipate for several reasons. Most worrying is that the estimated initial impact could have been due to EDE or SDB. The techniques discussed in Sections 3.3 and 3.5 can alleviate some of these worries. Perhaps respondents forget a treatment that is not particularly salient or interesting; or maybe, for topics encountered frequently in daily life, there are other countervailing forces that dampen or even counter the treatment's effect. In principle, the persistence of treatment effects can be assessed using follow-up surveys, where members of the original sample are recontacted and asked questions related to the dependent variables of interest, without administering the treatment again.

The degree of persistence of an experiment will depend on its type. Priming treatments' effects which simply change the (momentary) accessibility or salience of some beliefs are likely to dissipate quickly. Factorial studies have a different goal (namely, understanding the marginal effect of one or several specific features on a certain outcome in complex issues) and, per se, do not try to convey new content to the respondent (unless combined with an information or pedagogical experiment). Persistence is thus perhaps most relevant for information and pedagogical experiments. In this case, we may expect the effect to persist more if the initial treatment is more powerful and interesting enough to the respondent and if it has wider applicability, that is, it provides content that is usable in more situations. A follow-up can be useful to test for persistence, but it will not easily uncover the reasons that a treatment effect persists or not.

Recontacting the same respondents can be challenging, and the success rates will depend on the platform you use. **Supplemental Table A-4** reviews recontact rates across different studies and platforms and the persistence of treatment effects. Recontact rates differ drastically depending on the survey channel and the time between the first survey and the follow-up. Typically, to maximize recontact rates, you should think about increasing the incentives and offering extra rewards for people to take your follow-up survey (many commercial survey companies will do this), as well as making it as short as possible.

While the follow-up survey design broadly follows the guidelines already described here, there are the following two issues to pay attention to.

- Questions. The questions that are asked to assess the persistence of the effect should be identical to the original ones to avoid measurement error. If respondents are aware of their purpose, this may recreate EDEs and produce biased estimates.[19]
- Differential attrition. Especially when the recontact rate is low, there could be differential selection in the follow-up, causing problems similar to differential attrition (see Section 2.5 for a discussion of attrition). The best practice tips and possible corrections discussed in **Supplemental Appendix A-1.5** also apply to follow-up surveys.

Supplemental Material >

5. CONCLUSION

Surveys offer a unique opportunity to dive into people's minds to better understand how they reason, the things they care about, and their preferences. To ensure high data quality and reliable results, it is key to focus on proper design, sampling, and analysis. This article offered some practical recommendations for each step of the survey process that can hopefully help researchers across different fields make use of this valuable approach to research.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I thank my students and coauthors who have worked with me to develop these methods and have taught me a lot. I thank Alberto Binetti, Martha Fiehn, and Francesco Nuzzi for excellent research assistance.

LITERATURE CITED

Alesina A, Ferroni MF, Stantcheva S. 2021. *Perceptions of racial gaps, their causes, and ways to reduce them*. NBER Work. Pap. 29245

Alesina A, Miano A, Stantcheva S. 2023. Immigration and redistribution. *Rev. Econ. Stud.* 90:1–39

Alesina A, Stantcheva S, Teso E. 2018. Intergenerational mobility and preferences for redistribution. *Am. Econ. Rev.* 108(2):521–54

Allcott H, Taubinsky D. 2015. Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. *Am. Econ. Rev.* 105(8):2501–38

Althaus SL, Kim YM. 2006. Priming effects in complex information environments: reassessing the impact of news discourse on presidential approval. *J. Politics* 68(4):960–76

Bansak K, Hainmueller J, Hopkins DJ, Yamamoto T. 2018. The number of choice tasks and survey satisficing in conjoint experiments. *Political Anal.* 26(1):112–19

Bargh JA, Chartrand TL. 2014. The mind in the middle: a practical guide to priming and automaticity research. In *Handbook of Research Methods in Social Psychology*, ed. H Reis, C Judd, pp. 311–44. Cambridge, UK: Cambridge Univ. Press. 2nd ed.

Berinsky AJ, Huber GA, Lenz GS. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Anal.* 20(3):351–68

[19] One can run an obfuscated follow-up [as Haaland & Roth (2020) do] to hide the connection with the main survey.

Billiet J, Davidov E. 2008. Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociol. Methods Res.* 36:542–62

Billiet JB, McClendon MJ. 2000. Modeling acquiescence in measurement models for two balanced sets of items. *Struct. Equ. Model.* 7(4):608–28

Bogner K, Landrock U. 2016. *Response biases in standardised surveys*. GESIS Surv. Guidel., Leibniz Inst. Soc. Sci., Mannheim, Ger. **https://www.gesis.org/fileadmin/upload/SDMwiki/BognerLandrock_Response_Biases_in_Standardised_Surveys.pdf**

Bottan NL, Perez-Truglia R. 2022. Choosing your pond: location choices and relative income. *Rev. Econ. Stat.* 104(5):1010–27

Brader T, Valentino NA, Suhay E. 2008. What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *Am. J. Political Sci.* 52(4):959–78

Brehm J. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: Univ. Mich. Press

Byrne BM. 1989. Multigroup comparisons and the assumption of equivalent construct validity across groups: methodological and substantive issues. *Multivar. Behav. Res.* 24(4):503–23

Byrne BM, Shavelson RJ, Muthén B. 1989. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol. Bull.* 105(3):456–66

Cheung GW, Rensvold RB. 2000. Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *J. Cross-Cult. Psychol.* 31(2):187–212

Clifford S, Sheagley G, Piston S. 2021. Increasing precision without altering treatment effects: repeated measures designs in survey experiments. *Am. Political Sci. Rev.* 115(3):1048–65

Cohn A, Engelmann J, Fehr E, Maréchal MA. 2015. Evidence for countercyclical risk aversion: an experiment with financial professionals. *Am. Econ. Rev.* 105(2):860–85

Cohn A, Maréchal MA. 2016. Priming in economics. *Curr. Opin. Psychol.* 12:17–21

Dafoe A, Zhang B, Caughey D. 2018. Information equivalence in survey experiments. *Political Anal.* 26(4):399–416

De Jong MG, Steenkamp JBE, Fox JP, Baumgartner H. 2008. Using item response theory to measure extreme response style in marketing research: a global investigation. *J. Market. Res.* 45(1):104–15

Dechezleprêtre A, Fabre A, Kruse T, Planterose B, Sanchez Chico A, Stantcheva S. 2022. *Fighting climate change: international attitudes toward climate policies*. NBER Work. Pap. 30265

Dillman D, Smyth J, Christian L. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley

Epper T, Fehr E, Fehr-Duda H, Kreiner CT, Lassen DD, et al. 2020. Time discounting and wealth inequality. *Am. Econ. Rev.* 110(4):1177–205

Guiso L, Sapienza P, Zingales L. 2018. Time varying risk aversion. *J. Financ. Econ.* 128(3):403–21

Haaland I, Roth C. 2020. Labor market concerns and support for immigration. *J. Public Econ.* 191:104256

Haaland I, Roth C. 2023. Beliefs about racial discrimination and support for pro-Black policies. *Rev. Econ. Stat.* 105(1):40–53

Haaland I, Roth C, Wohlfart J. 2020. *Designing information provision experiments*. CESifo Work. Pap. 8406, CESifo, Munich, Ger.

Hainmueller J, Hangartner D, Yamamoto T. 2015. Validating vignette and conjoint survey experiments against real-world behavior. *PNAS* 112(8):2395–400

Hainmueller J, Hopkins DJ, Yamamoto T. 2014. Causal inference in conjoint analysis: understanding multidimensional choices via stated preference experiments. *Political Anal.* 22(1):1–30

Heen M, Lieberman JD, Meithe TD. 2020. *A comparison of different online sampling approaches for generating national samples*. Res. Brief, Cent. Crime Justice Policy, Univ. Nev., Las Vegas

Hvidberg KB, Kreiner C, Stantcheva S. 2021. *Social position and fairness views*. NBER Work. Pap. 28099

Israel A, Rosenboim M, Shavit T. 2014. Using priming manipulations to affect time preferences and risk aversion: an experimental study. *J. Behav. Exp. Econ.* 53:36–43

Johnson T, Kulesa P, Cho YI, Shavitt S. 2005. The relation between culture and response styles: evidence from 19 countries. *J. Cross-Cult. Psychol.* 36(2):264–77

Jöreskog KG. 2005. *Structural equation modeling with ordinal variables using LISREL*. Unpublished manuscript, Upssala Univ., Upssala, Swed.

Karadja M, Mollerstrom J, Seim D. 2017. Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution. *Rev. Econ. Stat.* 99(2):201–12

Krosnick JA. 1999. Survey research. *Annu. Rev. Psychol.* 50:537–67

Krosnick JA, Narayan S, Smith WR. 1996. Satisficing in surveys: initial evidence. *New Dir. Eval.* 70:29–44

Kuziemko I, Norton M, Saez E, Stantcheva S. 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. *Am. Econ. Rev.* 105(4):1478–508

Link S, Peichl A, Roth C, Wohlfart J. 2022. *Information frictions among firms and households*. CESifo Work. Pap. 8969, CESifo, Munich, Ger.

Merolla J, Ramakrishnan SK, Haynes C. 2013. "Illegal," "undocumented," or "unauthorized": equivalency frames, issue frames, and public opinion on immigration. *Perspect. Politics* 11(3):789–807

Morren M, Gelissen JP, Vermunt JK. 2011. Dealing with extreme response style in cross-cultural research: a restricted latent class factor analysis approach. *Sociol. Methodol.* 41(1):13–47

Pasek J, Krosnick J. 2010. Optimizing survey questionnaire design in political science: insights from psychology. In *The Oxford Handbook of American Elections and Political Behavior*, ed. JE Leighley, pp. 27–50. Oxford, UK: Oxford Univ. Press

Pew Res. Cent. 2022. Writing survey questions. *Pew Research Center*. Accessed Sept. 7, 2022. **https://www.pewresearch.org/our-methods/u-s-surveys/writing-survey-questions/**

Scheufele DA. 2000. Agenda-setting, priming, and framing revisited: another look at cognitive effects of political communication. *Mass Commun. Soc.* 3(2–3):297–316

Small ML. 2009. "How many cases do I need?" On science and the logic of case selection in field-based research. *Ethnography* 10(1):5–38

Stantcheva S. 2021. Understanding tax policy: How do people reason? *Q. J. Econ.* 136(4):2309–69

Stantcheva S. 2022. *Understanding of trade*. NBER Work. Pap. 30040

Steiner P, Atzmüller C, Su D. 2017. Designing valid and reliable vignette experiments for survey research: a case study on the fair gender income gap. *J. Methods Measur. Soc. Sci.* 7(2):52–94

US Census Bur. 2019. *Counting the hard to count in a census: select topics in international censuses*. Tech. Note, US Census Bur., Washington, DC

Weber M, D'Acunto F, Gorodnichenko Y, Coibion O. 2022. The subjective inflation expectations of households and firms: measurement, determinants, and implications. *J. Econ. Perspect.* 36(3):157–84

Welkenhuysen-Gybels J, Billiet J, Cambré B. 2003. Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *J. Cross-Cult. Psychol.* 34(6):702–22