SURVEYS · TYPES AND APPLICATIONS

Survey Experiments as Vehicles for Causal Knowledge Generation

A survey experiment randomly assigns respondents to different versions of a question, message or scenario. This article explains how the design produces causal knowledge, its main forms, its limits and how to use it well.

ModQu research article · 706 words · 5 references (APA 7th edition)

Most surveys describe. They tell us what share of customers are satisfied or how opinions differ between groups. A **survey experiment** goes further: it randomly assigns respondents to different versions of a question, message or scenario and compares the answers. Because chance alone decides who sees which version, any systematic difference between the groups can be attributed to the version itself. The survey becomes a tool for answering "what causes what?" as well as "what is the case?" (Mutz, 2011).

Why randomisation gives causal answers

In ordinary survey analysis, a link between two variables may reflect a third factor. Customers who saw an advertisement may like a brand more because the advertisement was shown to people who already liked it. Random assignment breaks such links. On average, the groups are alike in every respect, measured or not, except for the version they received. The difference in their answers is therefore an unbiased estimate of the effect of the treatment for the population the sample represents (Gaines et al., 2007).

When the sample is a probability sample of a population, the survey experiment combines two strengths that rarely meet. Laboratory experiments offer strong causal inference but usually on small, unrepresentative groups. Surveys offer representative samples but weak causal inference. A population-based survey experiment offers both (Mutz, 2011).

Common forms

- **Split-ballot wording experiments.** Half the sample receives one wording and half another. Rasinski (1989) found that support for public spending in the United States was much lower when the question referred to "welfare" than when it referred to "assistance to the poor".
- **Vignette experiments.** Respondents judge short scenarios in which features, such as a person's age or a product's price, vary at random.
- **Conjoint experiments.** Respondents choose between profiles that vary on several attributes at once, so the effect of each attribute can be estimated separately (Hainmueller et al., 2014).
- **Information and message tests.** One group receives a fact or a message and the other does not, showing how information changes views.

Figure 1 shows a hypothetical message test.

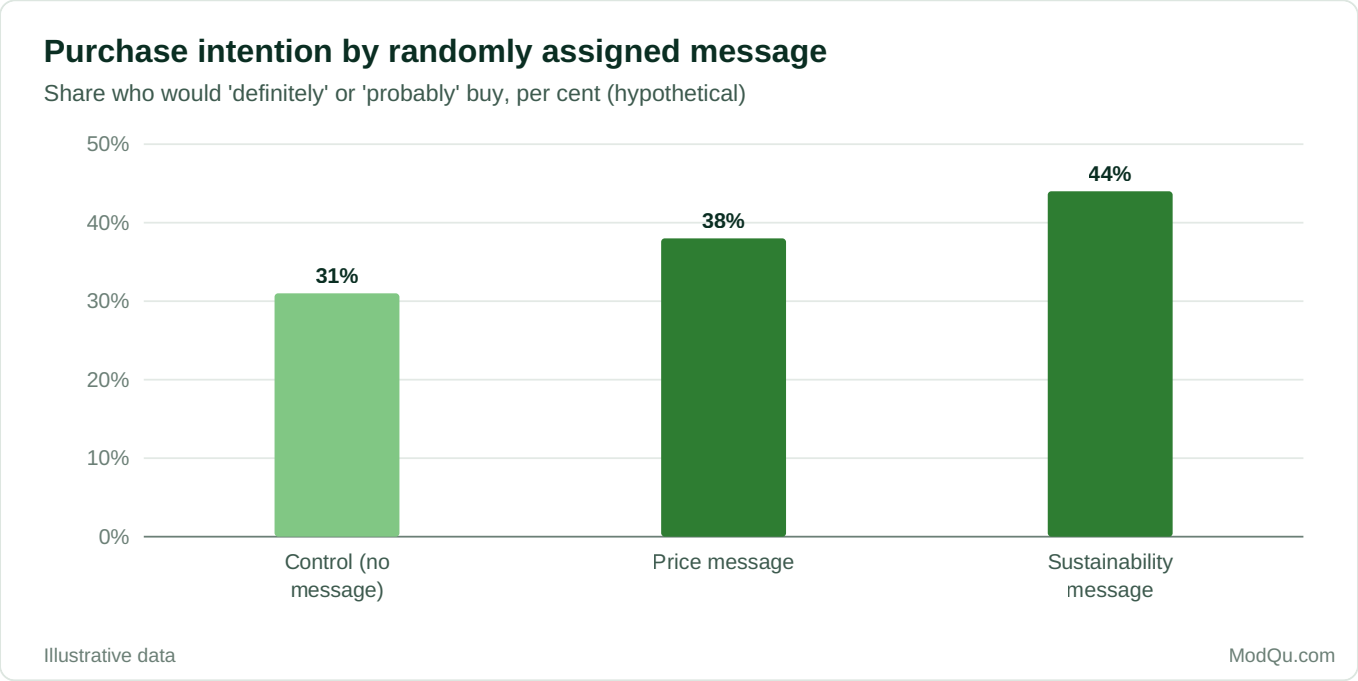


Figure 1. A hypothetical randomised message test. Each bar is a randomly assigned group of about 500 respondents. Illustrative data; chart by ModQu.

Uses in business research

Businesses use survey experiments to test product names, prices, package designs, advertising claims and the wording of customer communications before money is spent on a launch. Because the comparison is built into a single survey, the test is quick and inexpensive, and several versions can be compared at once. Analysis is usually simple: the difference in average answers or proportions between groups, with a confidence interval. Regression can add precision by adjusting for characteristics measured before the treatment, but randomisation, not the model, carries the causal claim.

What survey experiments can and cannot show

The causal claim holds only for the treatment as delivered in the survey. A short message read on a screen is not the same as an advertisement seen many times in daily life. Effects measured in a survey may be larger or smaller than effects in the real world, and stated intentions are not purchases. Gaines et al. (2007) also warned that earlier questions can "pre-treat" respondents, and that effects can fade quickly after the survey.

Generalisation is a second issue. Mullinix et al. (2015) replicated many experiments on both population samples and inexpensive online convenience samples and found that the treatment effects were broadly similar. This is encouraging, but it does not guarantee that every effect generalises, especially when the effect differs between types of people.

Good practice

1. State the hypothesis and the comparison before fielding the survey.
2. Randomise within the survey software and record which version each respondent saw.
3. Check that the groups are balanced on key characteristics.
4. Make sure each group is large enough to detect the effect that matters to the decision.
5. Include a manipulation check to confirm that respondents noticed the treatment.
6. Report the full wording of every version and all outcomes measured, not only the significant ones.

Key takeaways

A survey experiment turns a questionnaire into a source of causal knowledge at modest cost. Random assignment answers whether a wording, message or attribute changes responses. Interpreting what that means outside the survey still requires judgement about realism, timing and the population studied.

References

Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The logic of the survey experiment reexamined. *Political Analysis*, *15*(1), 1–20. https://doi.org/10.1093/pan/mpl008

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, *22*(1), 1–30. https://doi.org/10.1093/pan/mpt024

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly, 53*(3), 388–394. https://doi.org/10.1086/269158