

An Introduction to Survey Experiments Promises and Pitfalls

Yimeng Li

California Institute of Technology

Feb 10, 2021

Survey Experiment

- Social science researchers seek to establish causal relationships that are generalizable:
 - Lab experiment
 - Field experiment
 - Natural experiment
 - Survey experiment
- Examples of survey experiments in three subfields of political science:
 - American politics: Nelson, Clawson, and Oxley (1997)
 - International relations: Tomz and Weeks (2013)
 - Comparative politics: AARoe and Petersen (2014)

AP Example: Nelson, Clawson, and Oxley (1997)

Free Speech

Headline: Ku Klux Klan Tests OSU's Commitment to Free Speech

How far is OSU prepared to go to protect freedom of speech? The Ku Klux Klan has requested a permit to conduct a speech and rally on the Ohio State campus during the Winter Quarter of 1997. Officials and administrators will decide whether to approve or deny the request in December.

Numerous courts have ruled that the U.S. Constitution ensures that the Klan has the right to speak and hold rallies on public grounds, and that individuals have the right to hear the Klan's message if they are interested. Many of the Klan's appearances around Ohio have been marked by violent clashes between Klan supporters and counterdemonstrators who show up to protest the Klan's racist activities. In one confrontation last October in Chillicothe, Ohio, several bystanders were injured by rocks thrown by Klan supporters and protesters. Usually, a large police force is needed to control the crowds.

Opinion about the speech and rally is mixed. **Many students, faculty, and staff worry about the rally, but support the group's right to speak. Clifford Strong, a professor in the law school, remarked, "I hate the Klan, but they have the right to speak, and people have the right to hear them if they want to. We may have some concerns about the rally, but the right to speak and hear what you want takes precedence over our fears about what could happen."**

Public Order

Headline: Possible Ku Klux Klan Rally Raises Safety Concerns

Can campus police prevent a riot if the KKK comes to town? The Ku Klux Klan has requested a permit to conduct a speech and rally on the Ohio State campus during the Winter Quarter of 1997. Officials and administrators will decide whether to approve or deny the request in December.

Numerous courts have ruled that the U.S. Constitution ensures that the Klan has the right to speak and hold rallies on public grounds, and that individuals have the right to hear the Klan's message if they are interested. Many of the Klan's appearances around Ohio have been marked by violent clashes between Klan supporters and counterdemonstrators who show up to protest the Klan's racist activities. In one confrontation last October in Chillicothe, Ohio, several bystanders were injured by rocks thrown by Klan supporters and protesters. Usually, a large police force is needed to control the crowds.

Opinion about the speech and rally is mixed. **Many students, faculty, and staff have expressed great concern about campus safety and security during a Klan rally. Clifford Strong, a professor in the law school, remarked, "Freedom of speech is important, but so is the safety of the OSU community and the security of our campus. Considering the violence at past KKK rallies, I don't think the University has an obligation to allow this to go on. Safety must be our top priority."**

Note: Material that varied between the phrases is displayed in boldface. This material appeared in normal type in the experiment.

AP Example: Nelson, Clawson, and Oxley (1997)

- Outcome of interest:
 - Tolerance: Do you think that O.S.U. should or should not allow the Ku Klux Klan to hold a rally on campus?
 - Importance Rating
 - Importance Ranking

TABLE 5. Tolerance, Importance Ratings, and Importance Rankings by Framing Condition, Study 2

	Free Speech Framing Condition	Public Order Framing Condition
Tolerance ^a	4.38	3.54
Importance rating of free speech ^b	6.38	5.88
Importance rating of campus safety and security ^b	6.46	6.92
Importance ranking of free speech ^b	3.11	2.69
Importance ranking of campus safety and security ^b	2.84	3.00

^aHigher numbers indicate greater tolerance.

^bHigher numbers indicate greater importance.

IR Example: Tomz and Weeks (2013)

{Screen 1}

There is much concern these days about the spread of nuclear weapons. We are going to describe a situation the United States could face in the future. For scientific validity the situation is general, and is not about a specific country in the news today. Some parts of the description may strike you as important; other parts may seem unimportant.

Please read the details very carefully. After describing the situation, we will ask your opinion about a policy option.

IR Example: Tomz and Weeks (2013)

{Screen 2}

Here is the situation:

- A country is developing nuclear weapons and will have its first nuclear bomb within six months. The country could then use its missiles to launch nuclear attacks against any country in the world.
- The country [ally] signed a military alliance with the United States.
- The country [trade] high levels of trade with the United States.
- The country [regime].
- The country's nonnuclear military forces are half as strong as U.S. nonnuclear forces.
- The country's motives remain unclear, but if it builds nuclear weapons, it will have the power to blackmail or destroy other countries.
- The country has refused all requests to stop its nuclear weapons program.

By attacking the country's nuclear development sites now, the United States could prevent the country from making any nuclear weapons. Would you favor or oppose using the U.S. military to attack the country's nuclear development sites?

1. Favor strongly
2. Favor somewhat
3. Neither favor nor oppose
4. Oppose somewhat
5. Oppose strongly

IR Example: Tomz and Weeks (2013)

TABLE 1. The Effect of Democracy on Willingness to Strike

	United Kingdom (between)	United States (between)	United States (within)
Not a democracy	34.2	53.3	50.0
Democracy	20.9	41.9	38.5
Effect of democracy	-13.3	-11.4	-11.5
95% C.I.	(-19.6 to -6.9)	(-17.0 to -5.9)	(-14.7 to -8.3)

Note: The table gives the percentage of respondents who supported military strikes when the target was a democracy and when it was not. The difference is the estimated effect of democracy. In the United Kingdom, we obtained between-subject estimates by comparing 364 cases in which the target was a democracy, versus 398 cases in which it was not a democracy. In the United States, we obtained between-subject estimates by comparing 639 cases in which the target was a democracy, versus 634 in which it was not. The United States within-subject estimates were based on 972 respondents, each of whom assessed two scenarios, one in which the target was a democracy and another in which the target was not a democracy. 95% confidence intervals appear in parentheses.

CP Example: AAroe and Petersen (2014)

- Three conditions:
 - “Recipient with No Cues” condition: Imagine a man who is currently on social welfare.
 - “Unlucky Recipient” condition: Imagine a man who is currently on social welfare. He has always had a regular job, but has now been the victim of a work-related injury. He is very motivated to get back to work again.
 - “Lazy Recipient” condition: Imagine a man who is currently on social welfare. He has never had a regular job, but he is fit and healthy. He is not motivated to get a job.
- Respondents were then asked:
 - To what extent do you disagree or agree that the eligibility requirements for social welfare should be tightened for persons like him?

CP Example: AAroe and Petersen (2014)

TABLE 1 Opposition to Social Welfare by Experimental Condition and Country

	“Recipient with No Cues” Condition (Model I)	“Lazy Recipient” Condition (Model II)	“Unlucky Recipient” Condition (Model III)
Constant	.63*** (.05)	.73*** (.05)	.41*** (.05)
Country (Denmark) ¹	-.09*** (.02)	.03 (.02)	-.01 (.02)
Female	.002 (.02)	.04 (.02)	-.03 (.02)
Age	-.001 (.001)	.001 (.001)	-.002* (.001)
Education	.01 (.04)	-.05 (.04)	.05 (.05)
R ²	.03	.02	.01
N	613	651	642

Note: Entries are unstandardized ordinary least squares (OLS) regression coefficients with standard errors in parentheses. The dependent variable is opposition to social welfare, with higher values indicating more support for tighter eligibility requirements (i.e., higher values indicate a more opposition to social welfare). ¹Reference category for country variable is “United States.” * $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed t-test.

Survey Experiment

- Some considerations:
 - External validity: Barabas and Jerit (2010)
 - Demand effects: Mummolo and Peterson (2018)
 - Weighting: Miratrix, Sekhon, Theodoridis and Campos (2018)
- Two important classes of survey experiments:
 - Conjoint experiment and factorial experiment
 - List experiment, randomized response technique, and endorsement experiment

Conjoint Experiment

- Conjoint experiments (Green and Rao 1971) allow researchers to study the independent effects on preferences of many features of complex, multidimensional objects:
 - political candidates
 - immigrant admissions
 - public policies

Example: Hainmueller, Hopkins, and Yamamoto (2014)

Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

Example: Hainmueller, Hopkins, and Yamamoto (2014)

	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	<input type="radio"/>	<input type="radio"/>

On a scale from 1 to 7, where 1 indicates that the United States should absolutely not admit the immigrant and 7 indicates that the United States should definitely admit the immigrant, how would you rate Immigrant 1?

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

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

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

Conjoint Experiment: Design

- Consider a random sample of N respondents drawn from a population of interest.
- Each respondent (indexed by $i \in \{1, \dots, N\}$) is presented with K choice (or rating) tasks.
- In each of her tasks the respondent chooses the most preferred (or rates each) of the J alternatives (“profiles”).
- Each profile is characterized by a set of L discretely valued attributes, or a treatment composed of L components.

Conjoint Experiment: Identification Assumptions

- Assumption 1: Stability and No Carryover Effects
 - Potential outcomes always take on the same value as long as all the profiles in the same choice task have identical sets of attributes.
- Assumption 2: No Profile-Order Effects
 - The ordering of profiles within a choice task does not affect responses.
- Assumption 3: Randomization of the Profiles
 - Potential outcomes are statistically independent of the profiles.

Conjoint Experiment: Average Treatment Effect (ATE)

- Let $Y_i(\mathbf{t})$ denote the J -dimensional vector of potential outcomes for respondent i in her k th choice task that would be observed when the respondent received the sequence of profile attributes represented by \mathbf{t} , with individual components $Y_{ij}(\mathbf{t})$.
- Unit treatment effect:

$$\pi_i(\mathbf{t}_1, \mathbf{t}_0) \equiv Y_i(\mathbf{t}_1) - Y_i(\mathbf{t}_0)$$

- Average treatment effect:

$$\bar{\pi}(\mathbf{t}_1, \mathbf{t}_0) \equiv \mathbb{E}[Y_i(\mathbf{t}_1) - Y_i(\mathbf{t}_0)]$$

Conjoint Experiment: Average Marginal Component Effect (AMCE)

- The marginal effect of attribute l averaged over the joint distribution of the remaining attributes:

$$\begin{aligned}\bar{\pi}_l(t_1, t_0, p(\mathbf{t})) &\equiv \mathbb{E}\left[Y_i(t_1, T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) - Y_i(t_0, T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) \mid (T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) \in \tilde{\mathcal{T}}\right] \\ &= \sum_{(t, \mathbf{t}) \in \tilde{\mathcal{T}}} \mathbb{E}\left[Y_i(t_1, t, \mathbf{t}) - Y_i(t_0, t, \mathbf{t}) \mid (T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) \in \tilde{\mathcal{T}}\right] \\ &\quad \times p\left(T_{ijk[-l]} = t, \mathbf{T}_{i[-l]k} = \mathbf{t} \mid (T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) \in \tilde{\mathcal{T}}\right),\end{aligned}\tag{4}$$

$$\begin{aligned}\hat{\pi}_l(t_1, t_0, p(\mathbf{t})) &= \sum_{(t, \mathbf{t}) \in \tilde{\mathcal{T}}} \left\{ \mathbb{E}\left[Y_{ijk} \mid T_{ijkl} = t_1, T_{ijk[-l]} = t, \mathbf{T}_{i[-l]k} = \mathbf{t}\right] \right. \\ &\quad \left. - \mathbb{E}\left[Y_{ijk} \mid T_{ijkl} = t_0, T_{ijk[-l]} = t, \mathbf{T}_{i[-l]k} = \mathbf{t}\right] \right\} \\ &\quad \times p\left(T_{ijk[-l]} = t, \mathbf{T}_{i[-l]k} = \mathbf{t} \mid (T_{ijk[-l]}, \mathbf{T}_{i[-l]k}) \in \tilde{\mathcal{T}}\right).\end{aligned}\tag{5}$$

A Social Reference Theory of Sensitivity Bias

- Blair, Coppock, and Moor (forthcoming) theorize that sensitivity bias occurs for a given respondent if and only if all four of the following elements are present:
 - 1 A social referent the respondent has in mind when considering how to respond to a survey question. A social referent could be the respondent themselves.
 - 2 A respondent perception that the social referent can infer the subject's response to the sensitive question either exactly or approximately.
 - 3 A respondent perception about what response (or nonresponse) the social referent prefers.
 - 4 A respondent perception that failing to provide the response preferred by the social referent would entail costs to themselves, other individuals, or groups.

Sensitivity Bias and Indirect Questioning Techniques

- Direct survey questions on these topics often lead to a substantial amount of underreporting and nonresponse.
- To reduce possible biases due to social desirability and missing data, researchers increasingly rely upon several indirect questioning techniques, such as
 - List experiment (a.k.a. the item count technique, Miller 1984)
 - Randomized response technique (Werner 1965)
 - Endorsement experiment (Bullock, Imai, and Shapiro 2011)
- Consider an example from Rosenfeld, Imai and Shapiro (2014).

Direct Question

- Did you vote YES or NO on the Personhood Initiative, which appeared on the November 2011 Mississippi General Election ballot?
 - Voted Yes
 - Voted No
 - Did not vote
 - Don't know

List Experiment

- Respondents are randomly assigned to a control or a treatment condition.
 - Control respondents see a list of items and the question asks them how many they would respond to in the affirmative.
 - Treated respondents see an otherwise identical question except for the addition of a sensitive item.

List Experiment - Control Group

- Here is a list of four things that some people have done and some people have not. Please listen to them and then tell me **HOW MANY** of them you have done in the past two years. Do not tell me which you have and have not done. Just tell me how many:
 - Discussed politics with family or friends;
 - Cast a ballot for Governor Phil Bryant;
 - Paid dues to a union;
 - Given money to a Tea Party candidate or organization.

How many of these things have you done in the past two years?

List Experiment - Treatment Group

- Here is a list of four things that some people have done and some people have not. Please listen to them and then tell me HOW MANY of them you have done in the past two years. Do not tell me which you have and have not done. Just tell me how many:
 - Discussed politics with family or friends;
 - Cast a ballot for Governor Phil Bryant;
 - Paid dues to a union;
 - Given money to a Tea Party candidate or organization;
 - Voted 'YES' on the 'Personhood' Initiative on the November 2011 Mississippi General Election ballot.

How many of these things have you done in the past two years?

List Experiment - Analysis

- Identification assumptions (Imai 2011, Blair and Imai 2012):

Assumption 1 (No design effect): The inclusion of the sensitive item has no effect on respondents' answers to control items.

Assumption 2 (No liars): Respondents give truthful answers for the sensitive item.

- Canonical difference-in-means estimator $\hat{\tau}_{DiM}$:

Average response by the treated – Average response by the control

- Under *no design effect*, $\hat{\tau}_{DiM}$ estimates the population proportion **giving affirmative answers** to the sensitive item.
- Under *no liars*, $\hat{\tau}_{DiM}$ also estimates the population proportion **possessing** the sensitive item.

Randomized Response Technique (Forced Response)

- Randomization determines whether a respondent truthfully answers the sensitive question or simply replies with a forced answer, 'yes' or 'no'.
- To answer this question, you will need a coin. Once you have found one, please toss the coin and note the results of that toss (head or tail) on a sheet of paper. Do not reveal to me whether your coin lands on head or tail. After you have recorded the result of your coin toss, just tell me you are ready, and we will begin.
- To ensure that your answer is confidential and known only to you, please answer 'yes' if either your second coin toss came up heads or you voted 'YES' on the 'Personhood' Initiative, which appeared on the November 2011 Mississippi General Election ballot.
 - Yes
 - No
 - Don't know

Randomized Response Technique: Analysis

- Let
 - Z_i represent the latent binary response to the sensitive question for respondent i ;
 - Y_i represent the observed response (1 for 'yes' and 0 for 'no');
 - p denote the probability that respondent i is forced to answer 'yes'.
- Full compliance and no liars imply the following equality,

$$Pr(Y_i = 1) = p + (1 - p) \cdot Pr(Z_i = 1).$$

- This allows to derive the probability that a respondent truthfully answers 'yes' to the sensitive question,

$$Pr(Z_i = 1) = \frac{Pr(Y_i = 1) - p}{1 - p}.$$

Endorsement Experiment

- Respondents are randomly assigned to a control or a treatment condition.
 - In the control group, respondents are asked to evaluate some relatively uncontroversial issue or object;
 - In the treatment group, that issue or object is associated with the sensitive item before being evaluated.

Endorsement Experiment: Control Group

- We'd like to get your overall opinion of some people in the news. As I read each name, please say if you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of each person.
- Phil Bryant, Governor of Mississippi?
 - Very favorable
 - Somewhat favorable
 - Don't know/no opinion
 - Somewhat unfavorable
 - Very unfavorable

Endorsement Experiment: Treatment Group

- We'd like to get your overall opinion of some people in the news. As I read each name, please say if you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of each person.
- Phil Bryant, Governor of Mississippi, who campaigned in favor of the 'Personhood' Initiative on the 2011 Mississippi General Election ballot?
 - Very favorable
 - Somewhat favorable
 - Don't know/no opinion
 - Somewhat unfavorable
 - Very unfavorable

Endorsement Experiment: Analysis

- Idea: If this endorsement increases the level of support for the policy, then we interpret this effect as evidence that a respondent holds a favorable view toward the actor who endorses the policy.
- Analysis of endorsement experiments requires item response theory, which is beyond the scope of this class (see Bullock, Imai, and Shapiro 2011).