How can we observe the “Real Intention” in public opinion?

Based on research example using List Experiment*

HATA Masaki † ‡

Last Update 2018/4/30

Abstract

We discuss introducing a method of survey experiment that combines social survey and experimental methods and provide some application examples using survey experiments. In recent years, studies using randomized experimental approaches have rapidly increased in empirical political science. In fact, in 2015, one quarter of the articles in American Political Science Review, one of the most respected journals in political science, used experimental methods. Therefore, we focus on public opinion research and explain the various advantages of conducting randomized controlled trials (RCTs) on large scale social surveys. Furthermore, we report the results of survey experiments using RCTs that author conducted in the past. There are a variety of designs in survey experiments (e.g., frame experiment, list experiment, conjoint experiment, vignette experiment), and this report focuses on the list experimental method. List experiments with the Item Count Technique are used to analyze citizens’ attitudes towards sensitive political issues.

Specifically, we report the results of list experiments on “Image about The Democratic Party Japan (DPJ as Minshinto) on Japanese public opinion” conducted in February 2017. The DPJ’s approval rate has been decreasing over the past several years, and many voters do not support the DPJ. Previous studies suggest one possible reason for this opinion is policy failure or serious intra-party conflict while DPJ has been in power. This research aims to examine the influence of conspiracy on the Internet such as “5-channel” or “Matome-saito” on voters’ political attitudes. Since this experiment was carried out to elucidate rationales explaining why many Japanese harbor negative attitudes towards the DPJ, we focus extreme opinions such as “DPJ attempts to sell out Japan”. Moreover, we will discuss some limitations of experimental research in political science.

1 Introduction

This paper starts by providing an overview of the experimental approach currently used in political science and then examines the approach efficacy by looking at the case of list experiments. In recent years, the use of the experimental approach has increased dramatically in empirical political science. The experimental approach has long been used in fields such as economics and psychology, but it did not really catch on in political science until the start of the twenty-first century. To what extent has experimental research increased? Iida (2017) examined changes in

---

* Working paper at International Conference on Multicultural Democracy, Session IV.
† The University of Kitakyushu, Faculty of Law, Assistant Professor.
‡ e-mail: hatamasaki@kitakyu-u.ac.jp
the number of studies using the experimental approach among articles published in the *American Political Science Review* (APSR), an influential political science journal, since the year 2000. The results can be seen in Table 1. The table shows that the number of experimental studies increased enormously from around 2003. Moreover, as many as one quarter of articles in the APSR in 2015 were experimental studies, indicating that the experimental approach has now earned the title of a “common approach.” Yet when it comes to empirical political science in Japan, it appears that the experimental approach is still understood as a “new approach.” For example, the Japanese Journal of Electoral Studies, published by the Japanese Association of Electoral Studies, had their first feature on empirical political science in 2014 (Taniguchi, 2014). It seems (also based on my own intuition) that this approach became popular in Japan about a decade after it did in the United States.

Furthermore, experiments are a category that includes a diverse range of methods, such as survey and field experiments that mainly involve voters and natural experiments that use observational data. Survey experiments in particular are easier to use than other experimental approaches as this type is easily implemented and the analytical results are readily understood. As such, the use of this experimental approach has also increased in Japan, mainly in the fields of electoral and public opinion studies that involve voters.

As seen above, experimental political science is also thriving in Japan. Despite that, experimental research in Japan has yet to yield many results and currently Japanese researchers cannot help but depend on insights gained abroad (especially from American political science) when referencing actual examples of such studies. As such, this paper draws attention to list experiments, a category of survey experiments that is especially pioneering, and introduces studies I conducted in the past that looked at public opinion in Japan. Moreover, this paper concludes by examining not only the effectiveness of these types of experiments, but also the challenges we face in experimental political science with regard to external validity and reproducibility.

<table>
<thead>
<tr>
<th>Year</th>
<th>Dash Line (%)</th>
<th>Solid Line (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2002</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>2003</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>2004</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>2005</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>2006</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>2007</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>2008</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

(Source: Iida, 2017:49)

*1) The dash line: articles using Experimental Method

*2) The solid line: articles using Formal Model
2 Variations in Survey Experiments

2.1 The Basis of Principle of Experiments

Randomization is the most crucial aspect of the experimental approach (not limited to survey experiments). Specifically, conditions other than the predictor variables (covariates) can be considered homogenous as long as subjects are randomly assigned to particular experimental groups (called random assignment). In other words, by treating multiple groups under homogenous conditions as so-called parallel worlds, the effect of a treatment (called average treatment effect [ATE]) can be estimated by comparing the treatment group’s response $E(Y_1)$ to the treatment with a control group’s response $E(Y_0)$ (Rubin, 1974). That is, ATE can be expressed as Formula 1.

$$\tau_{ATE} = E[\tau] = E(Y_1 - Y_0) \quad \cdots \text{Formula 1}$$

I also apply this principle to the survey experiments below. Specifically, we can analyze as causal effect the differences in effect between a group provided with treatment and one not provided with treatment by randomly dividing respondents to a general awareness survey into a treatment group and a control group. In recent years, it has become possible to guarantee even stricter random allotment through the use of online systems such as Qualtrics, a survey assistance form.

2.2 Two Reasons Why Survey Experiments are Useful in Political Science

This paper examines survey experiments in particular. Survey experiments refer to an experimental method that uses questions in a social survey (a so-called sampling or awareness survey) involving the general population. Considering that the subjects are “people in general,” this approach resembles the laboratory experiments frequently used in psychology and other fields, which involve the gathering of students (subjects) in a laboratory. Nonetheless, survey experiments differ in two respects in particular when compared to laboratory experiments. First, there is a difference in sample size. Laboratory experiments have physical constraints that keep samples small, with each experimental group containing no more than a few dozen subjects. By contrast, large samples can be secured for survey experiments using sampling surveys, with each group containing hundreds of subjects. Second, there is the issue of sampling bias. Laboratory experiments involve subjects from a particular category, such as students, whereas most survey experiments involve a diverse range of respondents in terms of sex, age, and occupation.

These features are also one reason survey experiments are frequently used in political science. It has been pointed out for some time, especially in the field of political behavior, that social attributes such as sex, age, and occupation have a strong influence on many outcome variables, such as political participation, party identification, and voting behavior (Lazarsfeld et al., 1944;}

---

1 Due to space limitations, this paper will only briefly touch on past studies and methodologies for natural and field experiments as well as RCT frameworks, such as “causal inference using potential outcomes.” Refer to Imai (2017) for the fundamental principles of RCT and to Gerber and Green (2012) and Dunning (2012) for field and natural experiments.
Campbell et al., 1960). This indicates that laboratory experiments involving only students or any other small portion of voters cannot control for the effects of social attributes and thus do not mix well with research designs that aim to generalize results. By contrast, sampling bias is less of an issue in survey experiments that employ random sampling, making it possible to verify the causal effect of factor x on political attitudes. Considering the variety of assumptions that belong to political science, a major advantage of survey experiments is their potential for securing broad statistical populations.

2.3 Survey Experiment Methods

Survey experiments are not a uniform approach, however include a diverse range of experimental methods depending on the research questions. I will introduce two types of experiments in detail below.

2.3.1 Frame Experiments

The basic model for survey experiments is the frame experiment. It is a method for verifying differences in responses (response variables) between experimental groups when provided with different information. Below, I present Hata (2016) as an example of a frame experiment.

![Flow of the Frame Experiment](image)

The experiment conducted by Hata (2016) consisted of randomly dividing respondents into three experimental groups, each provided with different information: “A lower voting turnout among young people will reduce the quality of democracy” (treatment group 1), “A lower voting turnout among young people will reduce the benefits enjoyed by young people” (treatment group 2), and no information (control group). They were then asked about whether they have an interest in politics (the experiment flow is shown in Figure 1). Hata (2016) also stratified the sample into a younger layer (ages 18-19) and an older layer (ages 20-69), analyzing the data according to six patterns of three experimental groups across two generations. The results for the younger layer showed that only the average score of treatment group 1 was statistically significantly higher than the average score of the control group, while those for the older layer...
showed that only the average score of treatment group 2 was statistically significantly higher than the average score of the control group. In other words, the results of this experiment suggested that there exist generational differences in political images and political views. In this way, one characteristic of frame experiments is that providing randomized experimental groups with different information allows us to analyze how much of an effect a particular piece of information can have on groups (public opinion), both specifically and individually.

### 2.3.2 Conjoint Experiments

Moving on, I would like to talk about conjoint experiments. Conjoint analysis has frequently been used in the fields of marketing and business administration for some time. It is a useful approach when disaggregating attributes (elements) that make up products and evaluating what attributes influence the response variables. For example, there was a study that analyzed to what extent attributes that make up a fruit and vegetable producing area (location, cultivation method, and price) decide the brand strength of a fruit or vegetable as well as clarified how foreign products can reduce brand strength (Oura et al., 2002).

Conjoint experiments can be understood as conjoint analysis expanded by the addition of an experimental component. In ordinary conjoint analysis, all attributes are shown and respondents are asked to weigh them against each other and select the best (worst) one. By contrast, in a conjoint experiment, respondents are asked to pick the package with the most desirable profile from a selection of two or three packaged products that each randomly show one of the attributes. The main point about conjoint experiments is that it is not the subjects that are randomized, but the experimental items (attributes). It is also conventional to have the subjects respond several times (about three to five times), thereby gathering data on a large number of patterns and verifying the independent effect of each attribute on the response variables.

![Fig. 2: Screen from the Conjoint Experiment](image)

I would like to briefly talk about conjoint experiments in political science by presenting one of my own studies. We conducted a conjoint experiment every day over a five-day period leading up to the Japanese lower house election in October 2017. The experiment involved about 80 students from the Faculty of Law at the University of Kitakyushu. We prepared two fictitious candidates for this experiment and asked the students to pick “the desired candidate”.

---

2 More precisely, the students with a student number ending in an even number were asked to pick “the
preset the candidate attributes, including age (35, 45, 55, 65, and 75), most prioritized policy (improving childcare support, strengthening military power, making education free, and revising the constitution), and party affiliation (Liberal Democratic Party, Constitutional Democratic Party, Party of Hope, Communist Party, and independent), and then used the php program to combine them at random. In short, by likening the candidates to “products” and framing their age, prioritized policy, and party affiliation as “product attributes,” we adapted a conjoint analysis for an analysis of candidate selection (Song and Zenkyo, 2017).

You can see the actual survey screen in Figure 2. As you can see, the respondents in the experiment were presented with two candidates of different ages, prioritized policies, and party affiliations, and were asked four times to select the candidate who is best overall. In other words, we conducted a multiple regression analysis of the attribute effects by pooling experimental data corresponding to 80 students*4rounds*4days = (the number of total of )1,280 (Hainmueller et al., 2014). The results of the analysis showed that candidates with the attributes of being 35 years old, who want to improve childcare support, make education free, and belong to the Liberal Democratic Party were the most likely to be picked. Moreover, when analyzing the data chronologically, we found that policy became more important than age as Election Day drew closer. That is, by conducting a conjoint experiment, you can disaggregate a comprehensive political environment and analyze in greater detail what conditions influence political decision-making and to what extent.

3 Examples of Studies Using List Experiments

3.1 List Experiment Outline

Lastly, I will explain list experiments, which are the focus of this paper. List experiments are an exceedingly “interesting” analytical method, despite being more limited than the two aforementioned methods in terms of analysis subjects. This method is used mainly when analyzing the public opinion response to sensitive contents.

This is a slight change of topic, but there is no guarantee that respondents in a general sampling survey will respond “truthfully” to all questions. For example, if there is a question about experiences of drug use, respondents may not truthfully respond “Yes” even if they have had such experiences. People are reluctant to respond in ways that go against socially desirable norms and sometimes provide responses that are “lies.” This psychological tendency is generally called Social Desirability Bias. This bias frequently comes into play especially with political questions. For example, let us say there is the standard question “Did you vote?” in a sampling survey. Most results of such surveys will indicate a voting turnout of close to 90%, but the actual rate is about 50–60%. That is, about 30% of respondents are lying about having voted (Zenyko, 2016). In list experiments, attention is paid to differences in the “number of items” undesired candidate,” while those with an odd number were asked to pick “the desired candidate.” This also made it a field experiment for verifying the effect of negative recognition on voting turnout. Due to the intricacy of the data, here I only present the research design that involved students with odd numbers.

It would happen “by chance” that the candidates shown had the same attributes, in which case the respondents had been instructed beforehand to simply pick one and move on.

The response variables were binary, but Hainmueller et al. (2014) have shown that OLS regression (Ordinary Least Square) is not a problem when conducting a conjoint experiment.
for each randomized experimental group, making it a method of uncovering respondents “true sentiments” (Imai, 2011). I will explain this with reference to a specific example below.

3.2 Example: Is the DPJ as “Anti-Japanese”?  

I would like to present a list experiment that I conducted myself. I conducted this list experiment to analyze the relationship between online and offline public opinion in Japan. I conducted a sampling survey from February 13 to 15, 2017, involving 1,500 registered users of Yahoo Cloud Services. I briefly explain the background of the study below.

As is well-known, the Abe administration has been in place for six years. Other than positive causes such as the furtherance of an assertive national security policy and a positive assessment of Abenomics, a significant reason for such a lengthy administration is public distrust in opposition parties. In fact, the Democratic Party Japan(DPJ), while in power from 2009 to 2012, lost public support in one fell swoop due to various policy failures and a bad response to the Great East Japan Earthquake. This is evidenced by DPJ winning 308 seats in the 2009 lower house election, but only 57 in 2012, and their approval rating dropping to below 10%. Subsequently, they were more or less forced to dissolve the party in October 2017, splitting into the Party of Hope and the Constitutional Democratic Party. One reason for DPJ’s failure to gain public support is the various “groundless rumors” that can be seen online. For example, a typical case is the opinion that “DPJ is an ‘anti-Japanese party ’ submissive to South Korea and China.” Such views are commonly expressed online, but to what extent are they accepted by the “real public opinion”?

In general, we tend to regard approval of “conspiracy theories” like the one described above as socially reprehensible. That is, if we ask a question about this in a sampling survey, the majority of respondents are likely to respond “I do not agree” as a consequence of Social Desirability Bias. Yet those writing online are people too, so it is possible that many actually agree with such views. This is the issue that I examined with a list experiment.

3.3 Example: Explaining the Research Design

In a list experiment, respondents are provided with several items and patterns that include reprehensible content (treatment group) as well as patterns that do not include such content (control group). The respondents are then randomly allotted and asked to state the “number” of items with which they agree. For reference, I will explain the experiment in detail below.

The following shows a number of political opinions that can be found on TV, in newspapers, and online. Do you agree with any of them? Please respond with the number of opinions (items) with which you agree.

- Businesses that force employees to work long hours should be legally punished.

---

5 Yet as this experiment was not only about the Democratic Party Japan, but also included two other list experiments, the sample size might not match the plain distributions.

6 The order of the items for the experimental groups were also randomized.
We should consider postponing an increase of the consumption tax for any number of times until the economy is booming.

I would prefer that the state discontinued welfare policies such as pensions.

(Treatment: Many Diet members from DPJ are anti-Japanese.)

An actual screen from the experiment can be seen in Figure 3. The focus of this experiment was to determine how many respondents agreed with the opinion “Many Diet members from the Democratic Party are anti-Japanese.” The difference between the treatment group and the control group was whether this item was included or not. In other words, respondents “assigned to the control group” “by chance” were provided with a list of three items, namely, “long working hours,” “increased consumption tax,” and “abolishing pensions,” while respondents “assigned to the treatment group” “by chance” were provided with four items, having added “DPJ is anti-Japanese.” Respondents in each experimental group were asked to choose “the number of items with which they agree” (with a difference in the upper limit).

The list experiment was structured as follows. If the majority of respondents do not agree with the opinion “Many Diet members from DPJ are anti-Japanese,” then the average number of items with which respondents agree should not differ so much between the two experimental groups, since respondents in both the treatment group and the control group would actually be choosing from the three items “long working hours,” “increased consumption tax,” and “abolishing pensions.” Yet if there is a difference between the two groups’ averages, this would indicate that many picked “DPJ is anti-Japanese,” since its addition is the only difference between the two lists.

Moreover, list experiments are characterized by asking respondents to respond to number of items. With this research design, there is no need to express agreement with a particular opinion. Generally speaking, many would be against “abolishing pensions” and there should be some level of support for strictly penalizing long working hours and delaying an increase of the consumption tax. That is, the experiment is designed so that the average number for the control group should be around 1.5. It follows that even someone who agrees with “DPJ is anti-Japanese,” but
hesitates to respond accordingly, can respond with the quantities “1” or “2.” Such responses might still point to agreement with other items, thus creating a setup that better reveals the respondents’ true sentiments.

### 3.4 Example: Verifying the Experiment Results

Figure 4 shows the analytical results obtained on the basis of the abovementioned experimental design. First, let us verify the response distribution. Figure 4 shows that the treatment group’s distribution is heavier on the right side compared to that of the control group. That is, by adding information about “DPJ is anti-Japanese” in the treatment group, we have uncovered that a certain number of people agree with this opinion.\(^7\)

![Fig. 4: Response Distribution](image)

I then conducted a multiple regression analysis to examine the results. Figure 5 illustrates the results of the multiple regression analysis after inserting the number of items chosen in the experiment (min=1, max=4) as the response variable, and the experimental group binary variable (treatment group=1, control group=2) as the explanatory variable. Figure 5 also shows the analytical results for younger (39 years or less) and older people (40 years or above). In short, about 20% of respondents (public opinion) hold “in their minds” the conspiracy-theory-like view that “There are many anti-Japanese Diet members in DPJ.” When limiting the analysis to the older layer, the coefficient becomes even higher (0.26). These results suggest that negative information “of questionable veracity” about DPJ is also widespread in the real world.

### 4 Conclusion and Implications

This paper has examined the efficacy of experimental research in political science by explaining the background of the experimental approach as well as providing actual examples. I believe that list experiments in particular have the potential to broaden the range of political science research in the future. For example, voting behavior during the 2016 US presidential election and

\(^7\) Figure 4 only shows the key variables, but a total of six covariates, such as sex and age, were inserted for the analysis.
with regard to Brexit has garnered attention for completely overturning the results of sampling surveys conducted previously. These phenomena are very much thought to have been caused by Social Desirability Bias (Berinsky, 2017; Jolley et al., 2017). It is said that a dislike for political correctness has become widespread, and list experiments are an extremely useful method for analyzing such unseen changes in public opinion.

Nonetheless, there are a number of problems with survey experiments. One is that of external validity and reproducibility. Survey experiments are attractive to researchers who aim to explicate more generalizable mechanisms from the viewpoint of causal inference. Yet unlike economics and psychology, political attitudes tend to depend heavily on political context. As such, even though a certain result is obtained in one experiment, there is no guarantee that it can be reproduced in a different time and place. Accumulating insights from experimental studies can allow us to verify the robustness of those insights through meta-analysis. However, careful interpretation is also needed when conducting experimental political science research, which is still in its early stages.

Reference


(3998 words using LaTeX word count)