A Study on the Mental Foundations and Evolving Legal Norms Regarding Hate Speech in Japan: Bandwagon Effect or Social Desirability Bias

Yoshitaka Nishizawa
Doshisha University

nishizawa@mail.doshisha.ac.jp

5 October 2019

A paper presented at the Japanese Political Science Association Annual Meeting,
at Seikei University, Musashino, Tokyo, 5-6 October 2019

ABSTRACT

Using a web-based experimental survey specifically designed to access respondents' attitudes towards hate speech and governmental roles in regulating it, I address two questions: 1) why Japanese people are more supportive of governmental regulation of hate speech when "the challenged" are the target rather than when "*Zainichi* Koreans" are, and 2) whether or not people's attitudes towards hate speech are influenced by "social pressure." My analysis shows that the effect of "justifications" for regulating hate speech sometimes differ depending on the target. It also identifies some "bandwagon effects" on people's attitudes towards governmental regulation of hate speech.

1. Introduction

I am currently engaging, with six other scholars, in a project titled "A Study on the Mental Foundation and Evolving Legal Norm Regarding Hate Speech in Japan."[1] The overall objective of our study is to understand the mental foundations of hate speech in Japan, using a web-based experimental approach, to provide some scientific evidence to assist in evaluating Japan's current Hate Speech Prevention Law, formally titled "Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior against Persons Originating from Outside Japan," which was enacted in June 2016.[2]

Our attempt to study public attitude towards hate speech is not the first such effort in Japan. The Japanese government, for example, which has been conducting a series of public opinion surveys regarding human rights protection (Public Opinion Survey regarding Human Rights Protection), added for the first time a set of questions about hate speech in 2017.[3] But the questions were just ones among many that were directed towards other aspects of human rights issues in Japan, and therefore, the survey was not adequate to understand the complex nature of attitudes towards hate speech. This 2017 governmental survey caught much attention when it revealed that a majority (57%) of Japanese was aware of hate speech activities, including demonstrations, meetings and vans driven with agitating messages blaring through loudspeakers in public. We need, however, to be careful of these kinds of figures that are based on subjective responses. Because even if the respondents claimed "to have been aware" of such hate speech activities, it is well possible that the respondents never have actually witnessed the discriminatory expressions that aim to stir up violence against a specific minority group in the society (Kohno and Nishizawa 2019).

Our study, on the other hand, perhaps is one of the first attempts specifically designed solely about hate speech. It is also unique because it takes advantage of a web-based experimental approach. As I will explain later in more detail, by dividing our respondents into several

---

[1] The members include: Masaru Kohno and Kentaro Hirose (Waseda University), Ki'ichiro Arai (Tokyo Metropolitan University), Miwa Nakajo (Tsuda University), Go Murakami (Ritsumeikan University), and Hae Kim (Chiba University). Comments by Takeshi Iida (Doshisha University) and Tetsuya Matsubayashi (Osaka University) on my earlier drafts were instrumental. Asahi Obata and Kohei Yamamoto (Doshisha University) assisted me as research assistants. Editorial assistance by John McCall is appreciated. This work was supported by JSPS KAKENHI, Grant Number JP17KT0005.

[2] Please see "http://www.moj.go.jp/ENGLISH/m_jinken04_00001.html" and "http://www.moj.go.jp/content/001199550.pdf" (Accessed on 30 August 2019).

[3] The Japanese Cabinet Office has been regularly conducting similar surveys since 1958. URLs: "https://survey.gov-online.go.jp/h29/h29-jinken/index.html" and "http://www.moj.go.jp/content/001261929.pdf" (Accessed on 8 September 2019).

experimental groups and comparing their attitudes towards hate speech with those of the control groups, we can identify causal mechanisms of factors that may/may not foster supportive attitudes for governmental regulation against hate speech.

We have conducted such web-based experimental surveys twice already in March 2018 and March/April 2019. In both surveys, we designed the questionnaire aiming at the following objectives: They are 1) to identify who are more likely or less likely to support governmental regulation against hate speech, 2) to assess if their supportive attitude varies depending on targets (ie. victims of hate speech), 3) to examine if description of (or how we define) hate speech can influence respondents' attitudes and 4) to learn "justifications" for support of governmental regulations which may limit freedom of speech. In addition, for the March/April 2019 survey, we included experimental tools to test: 5) if social pressure (ie. "bandwagon effects" and "social desirability norm") can influence our respondents. Among these research questions, my present paper deals with the following two: "justifications" for supporting regulation (4), and possible effects of "social pressures" upon formation of attitudes towards hate speech (5).[4]

Let me at the outset present the results of my analysis. First, our four "justifications" for supportive attitudes towards regulation, including 1) "hate speech is offensive to the target groups," 2) "hate speech can harm the dignity of the victims," 3) "hate speech is unfair," and 4) "hate speech makes society worse," all do influence respondents' attitudes towards governmental regulation in a positive way as we expected. And, the effects seem to be stronger, for all "justifications" except for "offensiveness," among respondents in the "*Zainichi* Korean" experimental group than among those in "the challenged" experimental group (here, "challenged" includes people with some physical, intellectual, or mental disabilities).

Second, assuming our experimental stimuli work as we intend, (that is, assuming that we have been successful in "simulating" the social pressures, "bandwagon effects" and "social desirability norm," during our survey sessions) "bandwagon effects" seem to be operating, except for the "harm the dignity" variable, on the experimental groups whose target of hate speech is "*Zainichi* Koreans," but not operative for all "justification" variables for those whose target is "the challenged." The size of the effects, however, are rather small. Its substantive impact may not be large, but nevertheless people do "bandwagon" on the opinion of the majority.

[4] We are well aware of a major limitation of web-based experimental survey approach. Because our respondents are those who registered with a survey research company, they are not representative of our "population" (ie. Japanese citizens). What I present here, therefore, is by no means generalizable into the entire population.

The "social desirability norm," on the other hand, does not seem to influence attitudes for governmental regulation against hate speech in either of the experimental groups.

The rest of this paper proceeds in the following manner. In the next section, I will discuss some core arguments for possible "justifications" for governmental regulation towards hate speech. I will also review some literature about social pressures in public opinion formation. I will then explain our experiment design in section 3, and in section 4, I will present results of the data analysis. I will then conclude with some implications of this study.

2. The questions

There are two specific questions that I address in this paper.

The first is why Japanese people are more supportive of governmental regulation of hate speech when "the challenged" are the target rather than when "*Zainichi* Koreans" are the victims of such acts. With our 2018 survey, we discovered that, among three targets of hate speech, 1) "*Zainichi* Koreans," 2) challenged persons, and 3) minorities, the level of support for governmental regulation was higher when the target was "the challenged" than when it was for "*Zainichi* Koreans" (Kohno and Nishizawa 2019).[5] As you can see in Figure 1, with our 2019 survey, that tendency is once again observable. Along the scale of 0 to 6, 0 being "should not impose any restrictions" and 6 being "should impose thorough restrictions," mean scores of pro-regulation are 3.46 and 3.84 for "*Zainichi*" and " the challenged" experimental groups respectively.[6]

The second question concerns the effects of "social pressure" upon people's attitudes towards governmental regulation against hate speech. I consider here two types of social pressure, "bandwagon effects" and "social desirability norms." As I will discuss in more detail below, our basic assumption about people's attitudes towards hate speech is, to an extent, a product of social environment.

[5] In this case, "minorities" includes those people who are considered to be different from other people based on religious orientation, political belief, physical or mental difficulties, or sexual orientation.

[6] The t-test indicates that the difference is statistically significant with the probability that the difference being "by-change" is 0.000. The same tendency was confirmed in a pilot study conducted in 2016 by one of our research members, Masaru Kohno, with other comparative target groups, including "Brazilians who live in Japan" and "LGBT persons."

2.1 Possible "justifications" for supporting governmental regulation of hate speech

In America there is a strong belief, particularly among some constitutional scholars and political philosophers, that in order to protect free speech, even hate speech should not be regulated by law (Dworkin 2009, Strossen 2018, for example). While many of them acknowledge possible harms that hate speech can cause their targets, based on race, sexual preference, and religious belief, for example. Even so, democracy "requires" that everyone, including those who express hatreds, needs to have "a voice." Ronald Dworkin, for example, claims:

> "Fair democracy requires what we might call a democratic background: it requires, for example, that every competent adult have a vote in deciding what the majority's will is. And it requires, further, that each citizen have not just a vote but a voice: a majority decision is not fair unless everyone has had a fair opportunity to express his or her attitudes or opinions or fears or tastes or presuppositions or prejudices or ideals, not just in the hope of influencing others (though that hope is crucially important), but also just to confirm his or her standing as a responsible agent in, rather than a passive victim of, collective action (Dworkin 2009, vii)."

Some scholars, on the other hand, suggest several justifications for restricting hate speech. Bhikhu Parekh, for example, claims that hate speech is "unacceptable" for the following reasons:

> "Although free speech is an important value, it is not the only one. Human dignity, equality, freedom to live without harassment and intimidation, social harmony, mutual respect, and protection of one's good name and honor are also central to the good life and deserve to be safeguarded (Parekh 2009, 42/544 Kindle Edition)."

Although hate speech has long existed in Japan, as Kohno and Nishizawa (2019) point out, it caught public attention just after the Committee on Economic, Social and Cultural Rights of the UN Economic and Social Council made a specific recommendation on the hate speech situation in Japan to its government in May 2013.[7]

Because this is a relatively new issue, it is not clear if any Japanese are prepared to provide opinions about hate speech with regard to possible justifications for regulating such acts. We have, however, considered two sets of "justifications" in our experiment. They are "offensive

[7] Committee on Economic, Social and Cultural Rights of the UN Economic and Social Council, "Concluding observations on the third periodic report of Japan," adopted by the Committee at its fiftieth session (29April-17 May 2013).

nature" of hate speech and its possible "harm to the dignity" of an individual on one hand and the "unfair nature" of hate speech and its possibility of "making Japanese society worse," on the other.

The first set comes from the argument by Jeremy Waldron. Waldron is one of the leading scholars who calls for governmental regulation against hate speech.

He contrasts two logics of justifications, "protecting from offense" and "protecting dignity." Waldron insists that the former is not enough to justify laws restricting hate speech. For him offense is a "subjective reaction." Instead, he claims that the latter must be the objective of hate speech laws. According to him dignity refers "to their status as anyone's equal in the community they inhabit, to their entitlement to basic justice, and to the fundamentals of their reputation." For him dignity is a "public good" (Waldron 2012, 106-107).

Whether Japanese people can distinguish these two concepts is an empirical question. Furthermore, if these two logics of "justification" operate differently depending on the target of hate speech (ie. "*Zainichi* Koreans" or "the challenged"), that may explain, to some extent, why in Japan these two groups receive different levels of support for governmental regulation against hate speech targeting them.

The second set of "justifications," "unfair nature" of hate speech and its possibility of "making Japanese society worse," are drawn from recent efforts of the Japanese government to materialize the intent of the "Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior against Persons Originating from Outside Japan (hereafter "the Act"), which was enacted in 2016. Its basic stance towards hate speech clearly centers around these two ideas (ie. hate speech is "unfair" and it "makes society worse"). First, the Act does not use the word "hate speech;" rather, the Act refers it as "unfair discriminatory speech and behavior," as in the title and in the text of the Act. Second, the underlying assumption of the Act is that hate speech makes Japanese society worse. The materials the government prepared as a public relations campaign, such as posters and leaflets, do not directly use the phrase "making Japanese society worse," but instead they always include the following texts:

> "STOP! [sic] HATE SPEECH—Zero tolerance for hate speech" and
> "Let us acknowledge the differences among people and build together a society in which people mutually respect one's own human rights"[8]

[8] For the poster, please refer to the Appendix. It is accessible at: http://www.moj.go.jp/content/001184410.pdf. The translations are mine.

In short, we regulate hate speech, because it "makes Japanese society worse."

After two and a half years since the implementation of the Act, whether the Japanese people are aware of these two logics of "justification" for the Act, in itself is an interesting empirical question. Further, if these two logics operate differently depending on the target of hate speech, that may, once again, be the reason for the difference in the levels of support between the two target groups.

Thus, propositions to be tested here are:

> Proposition 1a ("Offensiveness proposition"): Respondents who think that insults and incitements to violence against [the target group] **would make them feel offended** are more likely to think that the government should impose restrictions on hate speech than those who think otherwise;
>
> Proposition 1b ("Dignity proposition"): Respondents who think that insults and incitements to violence against [the target group] **would harm their dignity** are more likely to think that the government should impose restrictions on hate speech than those who think otherwise;
>
> Proposition 1c ("Unfairness proposition"): Respondents who think that insults and incitements to violence against [the target group] **would be unfair** are more likely to think that the government should impose restrictions on hate speech than those who think otherwise;
>
> Proposition 1d ("Negative social effects proposition"): Respondents who think that insults and incitements to violence against [the target group] **would make Japanese society worse** are more likely to think that the government should impose restrictions on hate speech than those who think otherwise;

and

> Proposition 2 ("Differential effects proposition"): Some or all of the above "justification" factors (ie. Propositions 1a through 1d) operate differently depending on the target of hate speech, which leads to varying levels of support for governmental regulation against hate speech.

2.2 Effects of "social pressure"

People's attitudes towards hate speech depend on, on one hand, their individual characteristics and their own personal history; but, on the other hand, they are influenced by the context people live in. In our society people sometimes behave in accordance with the opinion of the majority, whether or not they feel its pressure. On the contrary, people sometimes behave in accordance

with the "expectation" that is believed to be "socially desirable" even when that "expectation" is in fact followed by only a small number of people. People's attitudes towards hate speech naturally are influenced by these "social pressures."

Among several possible kinds of "social pressure," in our web survey we considered "bandwagon effect" and the effect of "social desirability norm."

"Bandwagon effect" refers to "a situation where the information about majority opinion itself causes some people to adopt the majority view for whatever reason" (Marsh 1984, 51). Social scientists have discussed and studied it for a long time (Simon1957, Navazio 1977, Marsh 1984, Hardmeier 2008, and Murakami *et al*. 2016, for example). As Hardmeier summarizes, however, results of their cumulative efforts are "rather confusing as the findings are very disparate" (Hardmeier 2008, 2 of 14, Online edition).

According to Hardmeier there are at least five causal explanations for "bandwagon effects." They are: "contagion (emotional excitement or the enthusiasm of the crowds)," "gratification (supporting the winner)," "cue taking ('heuristic' information processing)," "cognitive-response (a process of self-persuasion)," and "strategic behavior (maximization of the utility of their action)" (7-9 of 14). Hardmeier's review is drown mostly from studies on voting behavior, particularly those that evaluate the effects on voters of public opinion polls published before elections.

The same logics may operate for people's attitudes towards hate speech. Before we can speculate about the mechanisms of "bandwagon effects" on hate speech, we first need to test if we can observe such effects. Thus, the propositions to be tested are:

> Proposition 3a ("Bandwagon effect proposition"): Because of the "bandwagon effect," respondents who are given the "majority support" scenario are more likely to support governmental regulation against hate speech than those who are given the "a little over half support" scenario, or than the control respondents who are given no such information.

Another type of "social pressure" has to do with "socially desirable norms." A human desire for impression management and self-presentation leads survey respondents to underreport their socially undesirable characteristics, attitudes, or behavior and to overreport what is considered socially "correct" or acceptable (Hyman 1957/1975, Maccoby and Maccoby 1954, Tourangeau *et al*. 2000, King and Bruner 2000, and Blinder *et al*. 2013). Blinder, Ford and Ivarsflaten (2013) argue that while many citizens "harbor" negative opinions and stereotypes about minorities,

including asylum seekers, immigrants, Muslims, and ethnic and racial minorities, they "respond to, and often internalize, a widespread social norm against prejudice and discrimination. (842)"

As I have mentioned earlier, the hate speech issue in Japan started to collect public attention recently, and it is not clear whether a "desirable norm" of how to deal with hate speech has been commonly shared in Japan. Given the Japanese government's recent efforts, however, to "educate" people about the hate speech issue, the "norm" may have started to sprout. Thus, the propositions to be tested are:

> Proposition 3b ("Social desirability proposition"): Because of the "social desirability bias" effect, respondents who are given the "socially desirable norm" scenario are more likely to support governmental regulation against hate speech than the control respondents who are given no such information.

3. Experiment design

Figure 2 represents our experimental model of study with four sets of variables. The dependent variable naturally is the attitude towards governmental regulation against hate speech. The independent variable is the treatment, whether a respondent is assigned to "*Zainichi*" target group or to "the challenged" group. There are four mediating variables which represent "justifications" to support/not support governmental regulations. They correspond to the four factors that I discussed in section 2.1. I will compare the effects of these mediating variables in each treatment set (ie. the target groups). Finally, there is a set of moderator variables. They represent factors that may influence relationships between the independent variables and the mediating variables and between the mediating variables and the dependent variable. They include respondents' demographic characteristics and their opinions about protection for freedom of speech from government regulation, for example.

Figure 3 illustrates the structure of the questionnaire for our 2019 web-survey, indicating the flow of questions and timing of random assignments for the different experimental groups. This structure was necessary to maximize the opportunity to test as many propositions as possible in one experiment. It is rather complicated. Let me, therefore, walk through the flowchart in the following section.

3.1 The first tier random assignment of two "target groups"

The questionnaire begins with an opening statement with a confirmation of consent, and that is followed by the first set of moderator questions. Then the respondents are randomly assigned into one of the two "Target" groups (ie. "*Zainichi* Korean" or "the challenged"). The "*Zainichi* Korean group" respondents, for example, read the following "Target Assignment Statement":

> "In present day Japan, you can find insults and incitements to violence against *Zainichi* Koreans in some areas or on the internet. Here, *Zainichi* Koreans means Korean nationals who reside in Japan (including special long-term residents who have lived in Japan before the Second World War, or their descendants).

whereas their "Challenged group" counterparts read:

> "In present day Japan, you can find insults and incitements to violence against those with disabilities in some areas or on the internet. Here, those with disabilities means people with some physical, intellectual, or mental disabilities.

3.2 The second tier random assignment of four experimental groups

For each of the above, the two first-tier groups, we once again randomly divided the respondents into four experimental groups.[9] We installed this tier of random assignment, to test propositions about the effect of "social pressure" upon attitudes towards governmental hate speech regulations, discussed in section 2.2 above. The four groups are the following:

1) Control groups;
2) Bandwagon 1, with "78% support" information groups;
3) Bandwagon 2, with "54% support" information groups; and
4) Social Desirability Norm groups

For the control group, the respondents received no additional information about the level of support among the general public for governmental regulation of hate speech.

We then simulated two "bandwagon" scenarios; one where a clear majority of people (78%) support the regulation and the other where just above half (54%) support it. The respondents in the first group, therefore read, right after the "Target Assignment Statement," the following information:

---

[9] Te be exact, there was another experimental group, "Abuse group," in which respondents read the following statement:

> "Do you think insults and incitements to violence such as "piss off", "You die", "Kill them" against *Zainichi* Koreans/those with disabilities would make them feel offended? Or do you think that it would not make them feel offended?

"By the way, last year we conducted a public opinion survey of voters in all prefectures in the country (some 3,000 people) and asked whether the government should regulate such activities. We obtained, among those who expressed a "yes/no" opinion to this question, the following result:

In favor of regulation: 78%
Against regulation: 22%

This result shows that there are quite a few people who think that it should be regulated. Keeping that in mind, please answer the following question."

The respondents in the second group read:

"By the way, last year we conducted a public opinion survey of voters in all prefectures in the country (some 3,000 people) and asked whether the government should regulate such activities. We obtained, among those who answered this question, the following result:

Support for regulation: 54%

This result shows that regulation is supported by a little over half of the people, and not necessarily by a great majority. Keeping that in mind, please answer the following question."[10]

Finally, we presented respondents in the "Social Desirability Norm" group, the following information:

"By the way, some people point out that such words and behaviors can create a sense of discrimination against the targeted groups. In order to foster a good society, "acknowledging differences among people and respecting each other" is considered desirable, if not the opinion of the majority in Japan. Keeping that in mind, please answer the following question."[11]

3.3 Mediator questions

Immediately after this "experimental group assignment," the four mediator questions follow in sequence. Please note that these mediator questions are "target-group-specific," meaning that the respondents are "forced" to answer these questions with respect to the target that they are assigned to. For example, the question about the offensive nature of hate speech reads, for the "*Zainichi* Korean group" respondents:

[10] The validity of this form of "stimulus" to simulate "bandwagon effects" has already been tested by some of our members (Murakami *et al*. 2016).

[11] Some of the wordings are based on the public relations poster prepared by the Japanese Ministry of Justice, also cited in Note 8 above.

> "Then do you think such insults and incitements to violence against *Zainichi* Koreans would make them feel offended? Or do you think that it would not make them feel offended?"

whereas for their "challenged group" counterparts, it reads:

> "Then do you think such insults and incitements to violence against those with disabilities would make them feel offended? Or do you think that it would not make them feel offended?"

This pattern remains the same for the remaining three mediator questions.

3.4 Attitude towards governmental regulation against hate speech, the main dependent variable

Finally, the main dependent variable question comes, and it reads, for the "*Zainichi* Korean group" respondents:

> "We ask you how the government should respond to this issue. Do you think that the government should impose restrictions on insults and incitements to violence against *Zainichi* Koreans? Or do you think that the government should not impose any restrictions?"

whereas for their "challenged group" counterparts, it reads:

> "Next, we ask you how the government should respond to this issue. Do you think that the government should impose restrictions on insults and incitements to violence against those with disabilities? Or do you think that the government should not impose any restrictions?"

Respondents answer this question by choosing one of the following response options.[12] The choices are:

0. Should not impose any restrictions;
1.
2.
3. Cannot say
4.
5.
6. Should impose thorough restrictions.

[12] To be exact, immediately before we asked this main dependent variable (ie. attitude towards governmental response to the issue) in order to ensure the "social pressure" remains in the mind of respondents, we made the respondents read the same "stimulus information." In other words, all experimental respondents read the "stimulus" information twice.

4. Data and results

To test the propositions introduced in section 2, I use the data from the web-based survey experiment conducted in March/April 2019.[13] We collected a total of 8,998 valid responses from pre-registered cyber panelists of one of the major survey companies in Tokyo.

In order to test the propositions, I estimated the following statistical model using a regression analysis:

$$\text{Support for Regulation} = b_1*\text{abu}[14] + b_2*\text{b78} + b_3*\text{b54} + b_4*\text{sdn} + b_5*\text{“justification”} + b_6*\text{ideology} + b_7*\text{free speech} + b_8*\text{female} + b_9*\text{age} + \text{constant}$$

The variables “b78,” “b54,” “sdn” are dummy variables for the three experimental groups, “bandwagon with 78% popular support,” “bandwagon with 54% popular support,” and “social desirability norm” respectively. The “justification” variable is a “surrogate” for the four justifications for support for governmental regulation. I alternate this “justification” variable with the four mediating variables. Therefore, I will have four sets of regression results. As moderator variables, I included “Ideology (0 to 10-liberal/conservative scale)” and opinion about the importance of protecting freedom of speech from government (4 point scale, “0” being “strongly disagree” to “1” being “strongly agree”). Finally, as control variables, a dummy variable for gender (“female,” “0” for male and “1” for female) and respondents’ age are included.

---

[13] We conducted two web-based experimental surveys in March 2018 and March/April 2019. For this paper I will solely rely on the second dataset, because only the latter had “stimuli” that simulate “social pressures.” In the dataset there are two waves of surveys. The first wave was conducted between Friday, 1 March and Friday, 15 March 2019 and the second was conducted between Wednesday, 10 April and Sunday, 14 April 2019. We decided to administer the second wave when we discovered that there was a possibility that we were not able to collect, at the end of the original survey period, enough sample for younger generations and for female samples. We used a different set of “cyber” samples for the second wave. For the final, combined distribution of samples by the key variables, please refer to Table 1. Numbers of respondents who completed the survey were 5,224 and 4,940 for March and April survey respectively, and numbers of valid samples after excluding “satisficers” based on two screening questions were 4,757 and 4,241, again respectively. We designed and programed the survey interfaces for both surveys using Qualtrix. Nikkei Research Inc. conducted and supervised the administration of both surveys.

[14] As explained in Note 9 above, there was another experimental group for “abuse.” Although I will not be discussing this variable in this paper, I included it into my analysis to take advantage of samples. The “asu” dummy picks any effects that its experimental “stimulus” can introduce, and therefore, the inclusion of these samples will not affect other estimates.

4.1 Estimation results

I estimated the above equation using a liner regression model. The results are summarized in Figures 4 through 11.[15]

4.1.1 "Offensiveness proposition"

Let us begin by evaluating the result for "Offensiveness proposition" (Figure 4). The figure is composed of two panels, one for "*Zainichi"* (top) and the other for "the challenged" (bottom) targets respectively. For each variable, the estimated coefficient is indicated by the dot, and the lines that extend on both sides of the dot show the 95% confidence interval for the corresponding estimate. The red dotted vertical line indicates "no effect." Therefore, if the confidence interval line runs across the red line, that variable is considered to be "statistically nonsignificant" at the 95% confidence level.

Please note that the distance between the red line and the dot depends on the unit of measurement for that particular variable. Therefore, absolute distances away from the red line cannot be compared between variables. For ease, particularly for variables whose estimated values are close to the red "no effect" line, and with short confidence interval lines, which makes it difficult to judge if the coefficients are statistically significant or not, I list in parentheses, on the side of the estimates, the probability of the variable being different from 0 by random chance. For example, in Figure 4, the probability for the first variable ("K-abu") is 0.06, which conventionally means that the variable is "not statistically significant" at 5% confidence level.

Let us now look at the effect of the "Offensiveness" variable. As you can see in Figure 4, the estimated coefficients for "Offensiveness," and their confidence intervals are clearly away from the red line, for both panels (ie. "*Zainichi*" and "the challenged" groups) which suggests that "Offensiveness proposition" is supported.[16] This effect can be observed again in Figure 5. In this figure, I overlaid two post-estimation simulation results in one chart. In a hypothetical situation in which all respondents think that hate speech "absolutely **would make** the victims feel offended" (ie. at the right end of the horizontal axis of the graph) their support for governmental regulation of hate speech increases by a little more than 2 points (to be exact, 2.5 and 2.2 for "*Zainichi"* and "the challenged," respectively), on a zero to six scale, as compared to

[15] To plot these summary chart of regression results, I used "ciplot" and "coefplot" modules with STSTA16.

[16] For the detailed estimation results, please consult Table 2.

another hypothetical situation in which all respondents think that hate speech "absolutely **would not make** the victims feel offended" (ie. at the left end of the axis).[17]

As for Proposition 2 for "Offensiveness," the "Differential effects proposition," does not seem to be supported by this analysis. As you can see in Figure 5, the two lines are almost perfectly parallel, indicating that the effect of "Offensiveness" does not differ for both "*Zainichi*" and "the challenged" groups. The difference between the two figures of "increase" discussed above (2.5 and 2.2) are not large enough to claim that they differ. I have conducted an independent "test for the difference in two coefficients," and it suggested that there is not a statistically significant difference between the two coefficients (Please see Table 2 for the figures. They are 2.536 and 2.241, for "*Zainichi*" and "the challenged" respectively.) (Colgg *et al*. 1995, Paternoster *et al*. 1998, Gelman and Stern 2006).

To sum up, Proposition 1a is supported, and therefore, the "offensive" nature of hate speech does influence peoples's attitudes towards governmental efforts to prevent acts of hate. People who think hate speech is "offensive" are more likely to welcome preventive efforts by the government. People who do not think hate speech is "offensive" are more likely to be cautious of governmental intervention. Proposition 2, however, is not supported, and therefore, the difference in the level of desirable governmental intervention between the "*Zainichi* Korean" target and "the challenged" target have nothing to do with the perceived level of "offensiveness" in hate speech itself. In other words, the difference in the level of desirable governmental intervention between the two group is caused by some other factors, but not by the "offensive" factor.

4.1.2 "Dignity proposition"

Figure 6 and 7 summarize the results for "Dignity proposition." The results are almost identical to the "Offensiveness" variable. As you can see in Figure 6, the estimated coefficients for "Harm Dignity," and their confidence intervals are clearly away from the red line for both panels (ie. "*Zainichi*" and "the challenged" samples), which suggests that "Dignity proposition" is supported. When people consider that hate speech would harm the dignity of the victims, they tend to welcome governmental intervention in preventing such acts.

---

[17] For this simulation, the values of other variables are fixed in their mean values. I used "margins" and "marginsplot" to produce the post-estimation simulation graphs and also used "combomarginsplot" to combine two "marginsplot" graphs for Stata. For those who are interested in the Stata program that produces the results presented in this paper, please visit my homepage at https://ynishiza.doshisha.ac.jp.

Figure 7 also suggests, as was the case for the "offensiveness" variable (ie. Figure 5), the increases in the level of support between the two ends of the horizontal axis, are 2.6 and 2.2 for "*Zainichi"* and "the challenged," respectively. It appears, therefore, that Proposition 1b, "Dignity proposition" is supported by our data.

Proposition 2 for "Harm Dignity," the "Differential effects proposition," may not seem to be supported because in Figure7 the two lines are almost parallel. The same statistical test, however, shows that the probability of the two coefficients being different by random chance is 0.02, so they are "statistically speaking" different.[18] The "Harm Dignity" justification, therefore, is a candidate for an answer to the puzzle: Because people associate the harmful nature of hate speech with their opinion about it in different ways, their levels of support for governmental regulation vary depending on the target.

4.1.3 "Unfairness proposition" and "Negative social effects proposition"

For both "Unfairness proposition" and "Negative social effects proposition," because their results of regression analyses exhibit a very similar pattern, let me summarize them together in this section.

Figures 8 and 9 summarize the results for the "Unfairness proposition." We can identify clear effects of "unfairness" variables on respondents' attitudes towards governmental regulation. Figures 10 and 11 summarize the results for the "Negative social effects proposition." The picture is the same. And, therefore, I can safely claim that Proposition 1c, "Unfairness proposition," and Proposition 1d, "Negative social effects proposition," are both supported by our data. Whether people consider that hate speech is "fair" or "unfair" clearly influences their acceptance of governmental involvement against such activities. When people think such acts are "unfair," they support governmental prevention efforts. At the same time, whether people think that hate speech makes Japanese society "better" or "worse" does seem to matter greatly. Those who think that such acts have negative influences on Japanese society naturally look for legal protection of the victims.

As for Proposition 2, the "Differential effects propositions" for "Unfairness" and "Negative social effects" appear to be supported. Figure 9 and 11 clearly illustrate these patterns. The statistical tests for the difference between two coefficients confirm that "Unfairness" and "Makes

[18] The coefficients are 2.564 and 2.236, between the two groups respectively (Table 3), with a difference of 0.328. The statistical test shows that the probability of the two coefficients being different by random chance is 0.0220.

Society Worse" variables are statistically different for "*Zainichi*" and "the challenged" samples, while probabilities for the coefficients being different by random chance are 0.000 for both sets.

In fact, the effects of "Unfairness" and "Makes Society Worse" variables for the "*Zainichi"* sample is strong enough to cancel the differences between the two target groups (Figures 9 through 11). We observe a relatively large difference in the attitudes at the left end of the scales (ie. when all respondents think that hate speech "absolutely would be fair," or "absolutely would make Japanese society better"). The differences, however, disappear at the right end of the scales (ie. when all respondents think that hate speech "absolutely would not be fair," or "absolutely would make Japanese society worse").

To sum up, both Proposition 1c, "Unfairness proposition" and Proposition 1d, "Negative social effects proposition" are supported. Proposition 2, "Differential effects proposition" for these two factors also is supported.

4.2 Role of "social pressure" on people's attitude towards governmental regulation against hate speech

Let me begin with the "bandwagon with 54% pro-support" scenario. None of the eight coefficients (4 mediating effects x 2 target groups, Figures 4, 6, 8 and 10, and Tables 2 through 5) for the "bandwagon with 54% pro-support" scenario (ie. "K-b54" and "C-b54" variables) showed statistically significant effects. It is understandable that the information of consent by merely a little over half of the general public would not create much significant "pressure." In theory, such information can work as "social pressure" only when respondents have a distorted image of society. They may have believed, before they took our survey, that a large majority of the population was supportive of governmental regulation. Or they may have believed that a large majority of the population opposed governmental regulation. Given that hate speech is a relatively new social issue, the chance is small that our respondents had such a distorted image.

Results for the "bandwagon with 78% pro-support" scenario is slightly positive. Each "K-b78" dummy variable for three "justification" variables, "Offensiveness," "Unfairness," and "Makes Society Worse" for the "*Zainichi*" samples, exhibits statistically significant effects at 5% confidence level. The sizes of the coefficients, however, are not large (minimum of 0.133 for "Offensiveness" in Table 3 and maximum of 0.167 for "Unfairness" in Table 5). In other words, their substantive influence on people's attitudes towards governmental regulation is limited at best.

Finally, let us look at the "Social desirability proposition." As summarized in Figures 4, 6, 8 and 10, and Tables 2 through 5, none of the eight coefficients for the dummy variables associated with the "social desirability norm" (ie. "K-sdn" and "C-sdn" variables) scenario showed statistically significant effects. I must, therefore, conclude Proposition 3b, "Social desirability proposition," was not supported by our data.

5. Conclusion

In 2016, the Japanese government enacted a hate speech law, formally titled "Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behavior against Persons Originating from Outside Japan," partly in response to pressure from the international community. The Act is a so-called "principle law" and does not criminalize hate speech. It only sets guidelines to follow for the national and local governments in eliminating hate speech in Japan. As Jyunko Kotani points out, however, the Act seems to have "been functioning in a way that controls the tone of hate speech expressed in the public sphere (Kotani 2017)."

My analysis showed that the "justifications" for governmental regulation of hate speech are clearly related to people's attitudes towards the regulation. In other words, their supportive/ unsupportive attitudes are not randomly formed. Therefore, any efforts to enlighten the general public, along the lines of the justifications I presented here, for example, would likely change the environment related to hate speech issue. My analysis also made it clear that not only does the level of support for governmental regulation vary depending on the target of hatred, but also the "effects" of the "justifications" vary, again, depending on the target. This suggests that any new attempts to extend the scope of protections against hate speech from "persons originating from outside Japan" to some other minority groups may need to consider a different approach for such attempts to be effective.

I could not elaborate, on the other hand, on reasons why the level of support for governmental regulation differs between the "*Zainichi* Koreans" and "the challenged," nor could I present reasons why the "effects" of "justification" differ depending on the target. I would like to propose to test some hypotheses for these puzzles in the next round of experimental surveys.

As for the effect of "social pressure," my analysis suggests that between "bandwagon effects" and effects of the "social desirability norm," only the former showed some influence, and only for the "*Zainichi* Korean" samples. In addition, the substantive size of its effect is small. It is not surprising, given that many previous studies also failed to identify such effects.

I would like to underscore, instead, that however limited the actual size of the effect, our simulation of "social pressure" did influence the attitudes of our respondents towards governmental regulation of hate speech. If, in fact, inserting one paragraph of information on a survey screen alluding to the fact that a majority of citizens do support governmental regulation can alter respondents' attitudes about hate speech, a bit of real "movement" in public opinion on this issue can create a bandwagon phenomenon. My research calls for careful attention to manipulative attempts by anybody, or any organizations for, or against, in realizing a society without hate speech.

Finally, I would like to point out two possible technical issues of our research design.

First, there is a possibility that our questionnaire format inflated the effect of "justification" variables in my regression analysis. As I explained in section 3 ("Experiment design"), our four "justification" questions were presented to our respondents right before our main variable, "support" or "do not support" governmental regulation of hate speech. It might have been difficult for respondents to say that the government "should not impose any restrictions on hate speech" right after answering that hate speech can harm the dignity of the targeted minorities, for example. It would have created psychological "dissonance" in their mind: hence, a higher correlation between the "justification" variables and my dependent variable. My next task would be to alter the sequence of questions to eliminate this possibility and re-evaluate the role of "justification" variables.

Second, there is a possibility that our "social desirability norm" stimulus has not been working as we intended. Although I did not report in the main text above, I have run the same regression analyses with the Japanese version of Balanced Inventory of Desirable Responding (BIDR-J) index in the equation (Tani 2008).[19] This index measures a tendency on the part of respondents to answer questions in the direction that "social desirable norm" points to. And, for all the equations I examined, except for one with "unfairness" for the "*Zainichi* Korean" sample, BIDR-J index turned out to be statistically significant, which suggests a possibility that our respondents were "influenced" by the "social desirability norm." The reason that our "sdn" dummy variable did not capture this tendency must be addressed in the future.

[19] Tani translated the BIDR index, developed by Delroy Paulhus, into Japanese. Paulhus' BIDR index is a shorter version of the original "Marlowe & Crowne Social Desirability Scale" (Crowne and Marlowe 1960, Paulhus 1991, and Tani 2008).

In any case, we are at a primitive stage of establishing a legal system to deal with hate speech in Japan. I hope this paper provides something useful in discussing this important and urgent issue of our society.

References

Blinder, Scott; Robert Ford; and Elisabeth Ivarsflaten. 2013. "The Better Angels of Our Nature: How the Antiprejudice Norm Affects Policy and Party Preferences in Great Britain and Germany." *American Journal of Political Science* 57 (4): 841 857.

Colgg, Clifford C.; Eva Petkova; and Adamantios Haritou. 1995. "Statistical Methods for Comparing Regression Coefficients Between Models." *American Journal of Sociology* 100: 1261-1293.

Crowne, D. P., & Marlowe, D. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24 (4): 349-354.

Dworkin, Ronald. 2009. "Foreword." In Hare, Ivan and James Weinstein, eds. *Extreme Speech and Democracy*. New York: Oxford University Press.

Gelman, Andrew and Hal Stern. 2006. "The Difference between 'Significant' and 'Not Significant' is not Itself Statistically Significant." *The American Statistician* 60 (4): 328-331.

Hardmeier, Sibylle. 2008. "The Effects of Published Polls on Citizens." In *The SAGE Handbook of Public Opinion Research*, edited by Wolfgnag Donsbach and Michael W. Traugott. London: Sage, 504-514.

Hyman, Herbert H. 1954/1975. *Interviewing in Social Research*. Chicago: University of Chicago Press.

King, Maryon F. and Gordon C. Bruner. 2000. "Social Desirability Bias: A Neglected Aspect of Validity Testing." *Psychology & Marketing* 17 (2): 79-103.

Kohno, Masaru and Yoshitaka Nishizawa. 2019. "Heito Supiichi Kisei eno Taido ha Dou Kimaruka." *Chuokoron* 133 (4): 166-180, ("How Attitudes towards Hate Speech are Formed")

Kotani, Jyunko. 2017. "A Comment on Hate Speech Regulation in Japan after the Enactment of the Hate Speech Elimination Act of 2016." *The Journal of law and politics, Shizuoka University* 21 (3&4) : 1 (228)-11 (218).

Maccoby, Eleanor E. and Nathan Maccoby.A 1954. "The Interview: A Tool of Social Science." In *Handbook of Social Psychology*, Vol. 1, edited by Gardiner Lindzey. Cambridge, MA: Addison-Wesley, 449-487.

Marsh, Catherine. 1985. "Back on the Bandwagon: The Effect of Opinion Polls on Public Opinion." *British Journal of Political Science* 15 (1): 51-74.

Murakami, Go; Kiichiro Arai; and Masaru Kohno. 2016. "Tasuha keisei ni kansuru chosa jikken." In *Jiken Seijigaku* edted by Yoichi Hizen. Tokyo: Keiso Shobo: 65-86 ("A Survey Experiment about Majority Opinion Formation" in *Experimental Political Science,* The chapter and the book titles are translated by Nishizawa).

Navazio, Robert. 1977. "An Experimental Approach to Bandwagon Research." *Public Opinion Quarterly* 41: 217-225.

Parekh, Bhikhu. 2009. "Is There a Case for Banning Hate Speech?" In Hare, Ivan and James Weinstein, eds. *Extreme Speech and Democracy*. New York: Oxford University Press..

Paternoster, Raymond; Robert Brame; Paul Mazerolle; and Alex Piquero. 1998. "Using the Correct Statistical Test for the Equality of Regression Coefficients." *Criminology* 36 (4): 859-866.

Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes*, edited by John P. Robinson, Phillip R. Shaver and Lawrence S. Wrightsman. San Diego: Academic Press, 17-59.

Simon, Herbert A. 1954. "Bandwagon and Underdog Effects and the Possibility of Election Predictions." *Public Opinion Quarterly* 18: 245-253.

Strossen, Nadine. 2018. *HATE: Why We Should Resist It with Free Speech, Not Censorship*. New York: Oxford University Press.

Tani, Iori. 2008. "Baransugata Syakaiteki Nozomashisa Han'noshakudo, Nihongoban (BIDR-J) no Sakusei to Shinraisei/Datousei no Kento" *Paasonariti Kenkyu* 17 (1): pp. 18-28 ("Creating the Balanced Inventory of Desirable Responding-Japanese Version and a study of its reliability and validity" in *the Japanese Journal of Personality*, The article title translated by Nishizawa)

Tourangeau, R., Rips, L., and Rasinski, K. 2000. *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.

Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, Massachusetts: Harvard University Press.

Tables and Figures

Table 1: Basic Statistics for Key Veriables

	Distribution		Support for Governmental Hate Speech Regulation	
	Frequencies	%	Means	S.D.
Target :				
Zainichi	4,469	49.67	3.459	1.447
The Challenged	4,529	50.33	3.841	1.302
Gender :				
Male	5,889	65.50	3.625	1.465
Female	3,102	34.50	3.701	1.226
Age Groups :				
20-29	662	7.36	3.578	1.383
30-39	1,002	11.14	3.468	1.412
40-49	1,947	21.64	3.532	1.377
50-59	3,003	33.37	3.644	1.395
60-69	2,384	26.49	3.854	1.359
Ideology :				
0. Liberal	99	1.16	4.163	2.148
1.	103	1.21	4.204	1.671
2.	281	3.30	4.192	1.468
3.	687	8.07	3.960	1.359
4.	948	11.14	3.808	1.270
5.	3,121	36.68	3.569	1.237
6.	1,412	16.59	3.634	1.250
7.	883	10.38	3.569	1.419
8.	605	7.11	3.511	1.641
9.	192	2.26	3.217	1.757
10. Conservative	178	2.09	2.807	2.224
Government Should Protect Freedom of Speech :				
Strongly disagree (0)	180	2.06	3.314	1.943
Somewhat disagree (.33)	1,548	17.72	3.498	1.429
Somewhat agree (.67)	4,192	47.99	3.639	1.254
Strongly agree (1)	2,816	32.23	3.789	1.513

Table 2: Support for Governmental Hate Speech Regulation,
Regression Analysis: Effect of "Offensiveness" by Target Groups

	Coefficients	Standard Errors	t	P > |t|
Zainichi Sample:				
K-abuse (K-abu)	.123	.065	1.90	0.058
K-bandwagon 1 (K-b78)	.133	.065	2.04	0.041
K-no band 2 (K-b54)	-.066	.065	-1.02	0.307
K-SD (K-sdn)	-.026	.065	-0.40	0.690
Offensiveness	2.536	.104	24.42	0.000
Ideology(0: Liberal - 10: Conservative)	-.139	.012	-11.97	0.000
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.146	.083	1.75	0.080
Female	.003	.045	0.07	0.947
Age	.003	.002	1.76	0.078
Constant	1.714	.162	10.60	0.000
(Adj R-squared: 0.183 Number of observation: 4,081)				
The Challenged Sample:				
C-abuse (C-abu)	-.015	.061	-0.25	0.804
C-bandwagon 1 (C-b78)	.060	.060	0.99	0.324
C-no band 2 (C-b54)	.000	.061	0.01	0.994
C-SD (C-sdn)	-.103	.060	-1.71	0.088
Offensiveness	2.241	.114	19.58	0.000
Ideology(0: Liberal - 10: Conservative)	-.024	.011	-2.25	0.025
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.106	.078	1.35	0.177
Female	-.002	.043	-0.04	0.969
Age	.008	.002	4.99	0.000
Constant	1.474	.156	9.44	0.000
(Adj R-squared: 0.101 Number of observation: 4,148)				

Table 3: Support for Governmental Hate Speech Regulation,
Regression Analysis: Effect of "Harm Dignity" by Target Groups

	Coefficients	Standard Errors	t	P > |t|
Zainichi Sample:				
K-abuse (K-abu)	.110	.064	1.71	0.087
K-bandwagon 1 (K-b78)	.125	.064	1.94	0.052
K-no band 2 (K-b54)	-.067	.064	-1.04	0.301
K-SD (K-sdn)	-.023	.064	-0.35	0.725
Harm Dignity	2.564	.095	27.02	0.000
Ideology(0: Liberal - 10: Conservative)	-.128	.012	-11.12	0.000
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.148	.082	1.81	0.071
Female	-.009	.044	-0.19	0.847
Age	.002	.002	1.31	0.190
Constant	1.735	.156	11.10	0.000
(Adj R-squared: 0.206 Number of observation: 4,077)				
The Challenged Sample:				
C-abuse (C-abu)	-.020	.061	-0.33	0.743
C-bandwagon 1 (C-b78)	.054	.060	0.90	0.369
C-no band 2 (C-b54)	-.009	.060	-0.14	0.886
C-SD (C-sdn)	-.100	.060	-1.67	0.095
Harm Dignity	2.236	.107	20.81	0.000
Ideology(0: Liberal - 10: Conservative)	-.023	.011	-2.14	0.032
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.095	.078	1.22	0.221
Female	-.003	.042	-0.08	0.936
Age	.008	.002	4.74	0.000
Constant	1.526	.152	10.06	0.000
(Adj R-squared: 0.110 Number of observation: 4,149)				

Table 4: Support for Governmental Hate Speech Regulation, Regression Analysis: Effect of "Unfairness" by Target Groups

	Coefficients	Standard Errors	t	P > |t|
Zainichi Sample:				
K-abuse (K-abu)	.163	.064	2.57	0.010
K-bandwagon 1 (K-b78)	.167	.064	2.61	0.009
K-no band 2 (K-b54)	-.051	.064	-0.80	0.426
K-SD (K-sdn)	.014	.064	0.22	0.825
Unfairness	2.13	.075	28.55	0.000
Ideology(0: Liberal - 10: Conservative)	-.117	.012	-10.21	0.000
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.179	.082	2.19	0.029
Female	.001	.044	0.02	0.981
Age	.004	.002	2.27	0.024
Constant	2.06	.149	13.85	0.000
(Adj R-squared: 0.219 Number of observation: 4,063)				
The Challenged Sample:				
C-abuse (C-abu)	.071	.061	1.15	0.250
C-bandwagon 1 (C-b78)	.048	.061	0.78	0.433
C-no band 2 (C-b54)	-.006	.061	-0.09	0.928
C-SD (C-sdn)	-.118	.061	-1.94	0.053
Unfairness	1.285	.075	17.07	0.000
Ideology(0: Liberal - 10: Conservative)	-.022	.011	-1.97	0.049
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.141	.079	1.78	0.075
Female	.035	.043	0.81	0.419
Age	.010	.002	5.73	0.000
Constant	2.28	.141	16.16	0.000
(Adj R-squared: 0.082 Number of observation: 4,139)				

Table 5: Support for Governmental Hate Speech Regulation, Regression Analysis: Effect of "Makes Society Worse" by Target Groups

	Coefficients	Standard Errors	t	P > |t|
Zainichi Sample:				
K-abuse (K-abu)	.125	.060	2.09	0.037
K-bandwagon 1 (K-b78)	.152	.060	2.53	0.012
K-no band 2 (K-b54)	-.035	.060	-0.58	0.561
K-SD (K-sdn)	.030	.060	0.50	0.620
Makes Society Worse	3.08	.081	38.09	0.000
Ideology(0: Liberal - 10: Conservative)	-.080	.011	-7.30	0.000
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.045	.077	0.59	0.558
Female	-.017	.041	-0.41	0.684
Age	.002	.002	1.14	0.255
Constant	1.41	.142	9.87	0.000
(Adj R-squared: 0.310 Number of observation: 4,064)				
The Challenged Sample:				
C-abuse (C-abu)	.011	.060	0.18	0.860
C-bandwagon 1 (C-b78)	.079	.059	1.33	0.183
C-no band 2 (C-b54)	.021	.059	0.35	0.729
C-SD (C-sdn)	-.083	.059	-1.40	0.161
Makes Society Worse	2.19	.092	23.69	0.000
Ideology(0: Liberal - 10: Conservative)	-.019	.011	-1.81	0.071
Government Should Protect Freedom of Speech (0: Strongly disagree - 1: Strongly agree)	.040	.077	0.52	0.603
Female	-.005	.042	-0.13	0.900
Age	.008	.002	4.82	0.000
Constant	1.68	.142	11.79	0.000
(Adj R-squared: 0.1348 Number of observation: 4,145)				

Figure 1: Support for Regulating HS by Target Groups

[N=8,774]

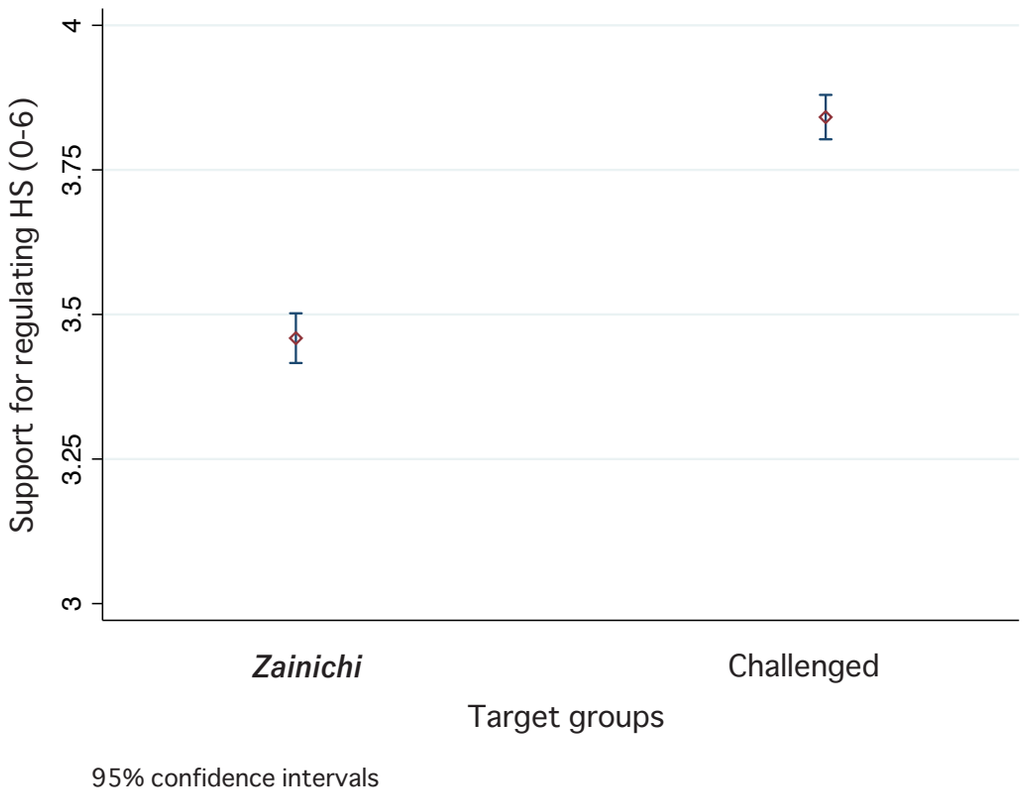


Figure 2: Experiment Model

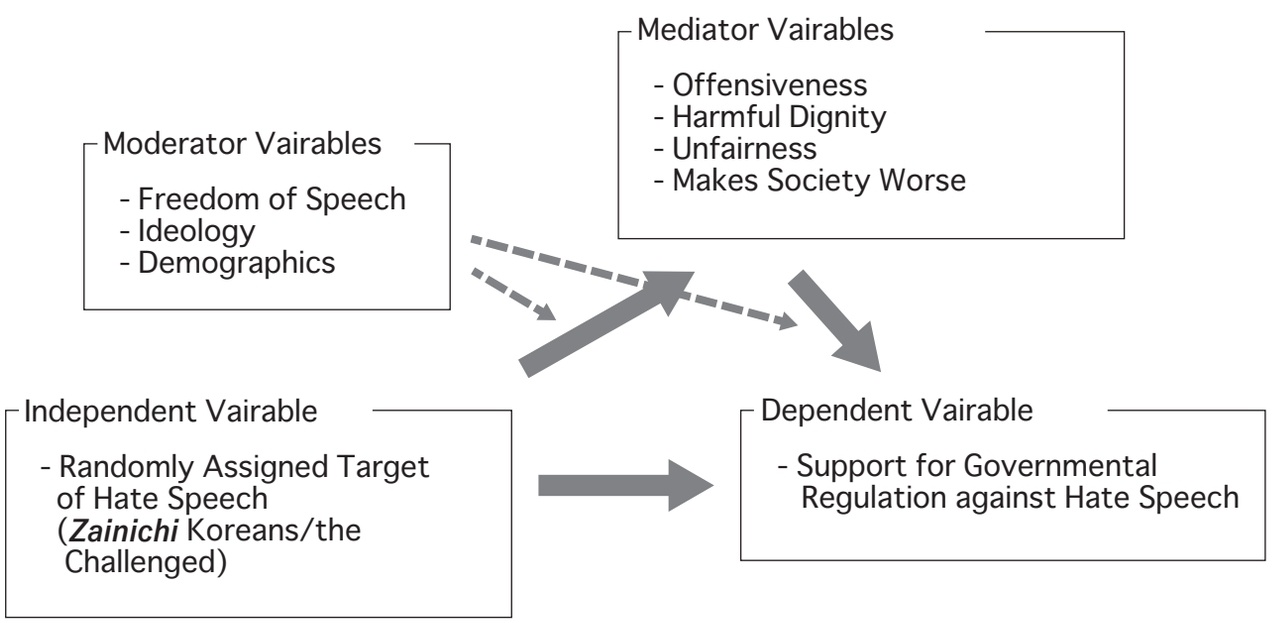


Figure 3: Flowchart of Survey Questionnaire, Ten Experimental Groups

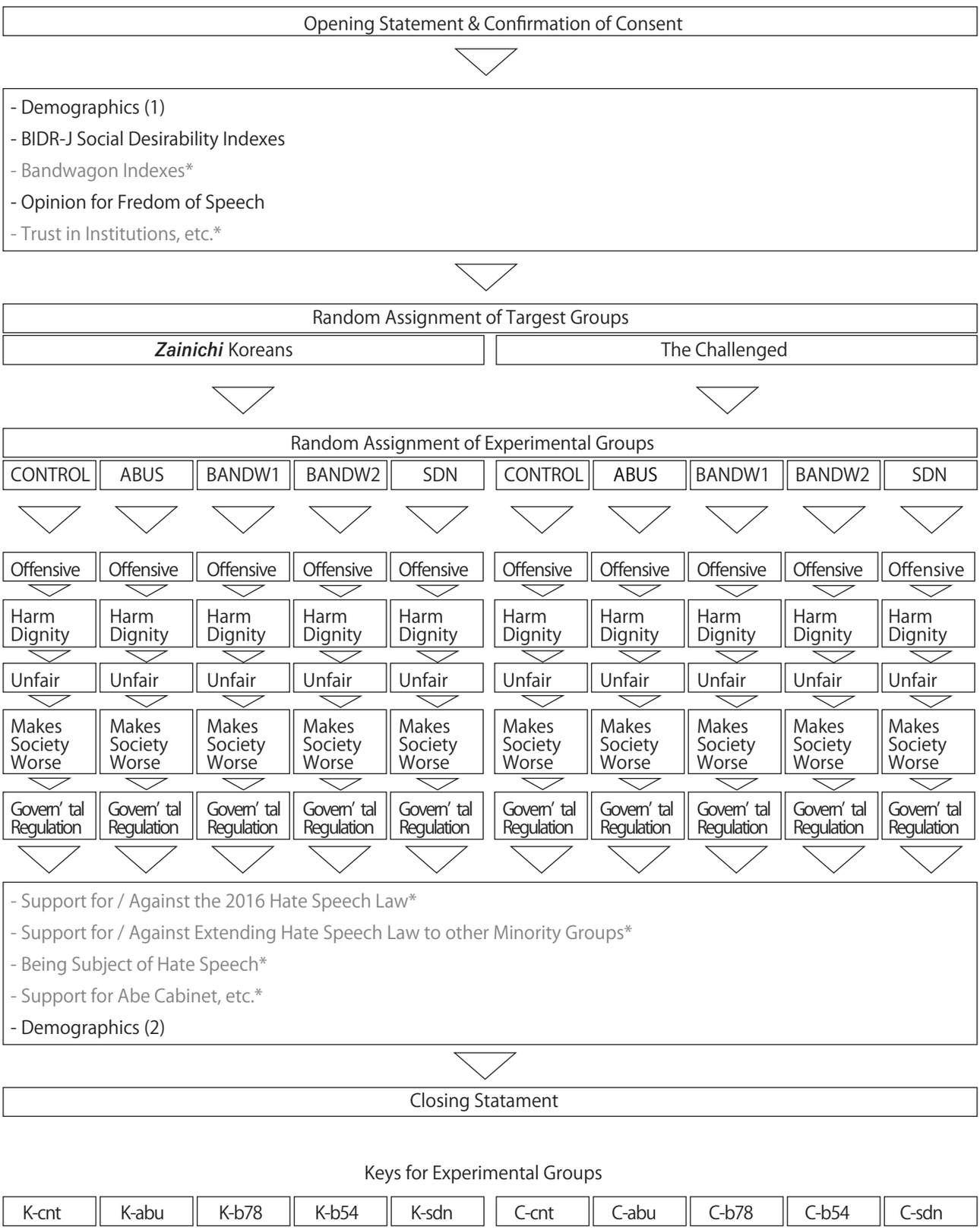


Figure 4: Support for Regulating HS -- Regression Analysis:
Effect of "Offensiveness" by Target Groups

[***Zainichi*** N=4,081; Challenged N=4,148]

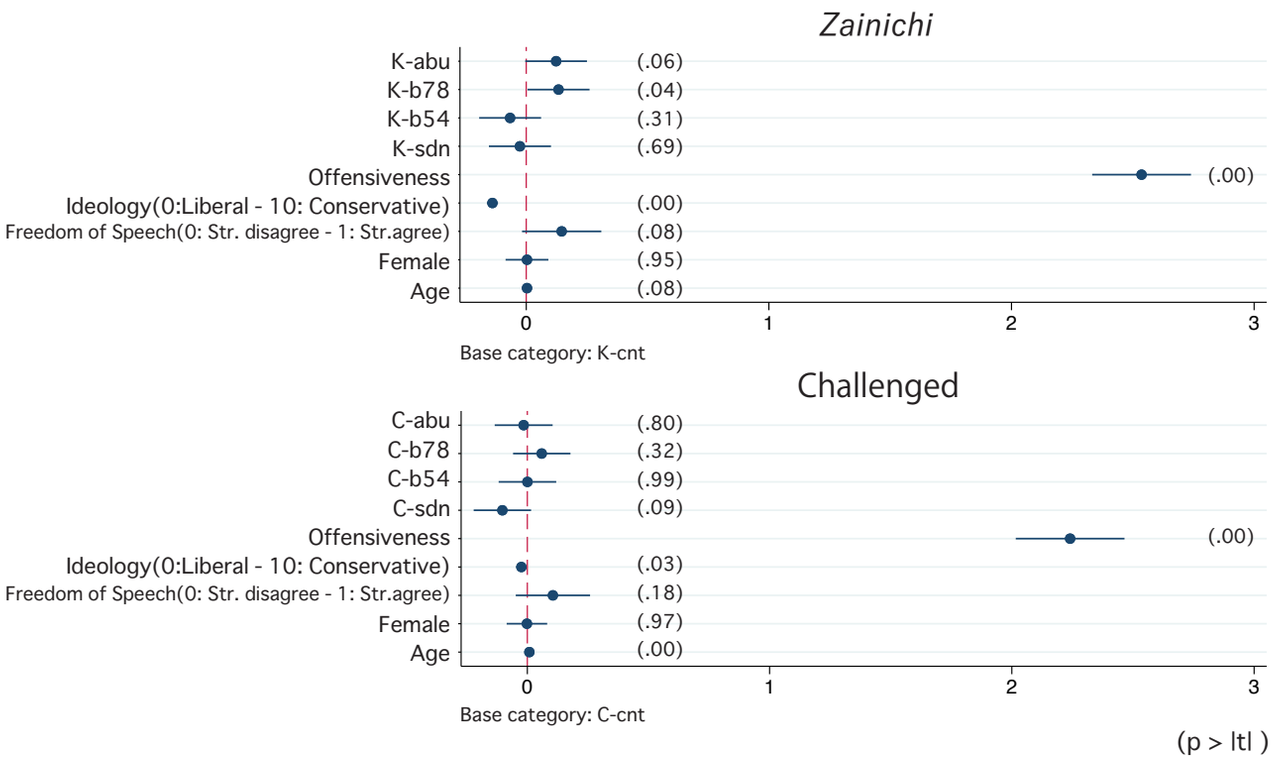


Figure 5: Support for Regulating HS with Bandwagon Effects
Post-estimation Simulation "Offensiveness" by Target Groups

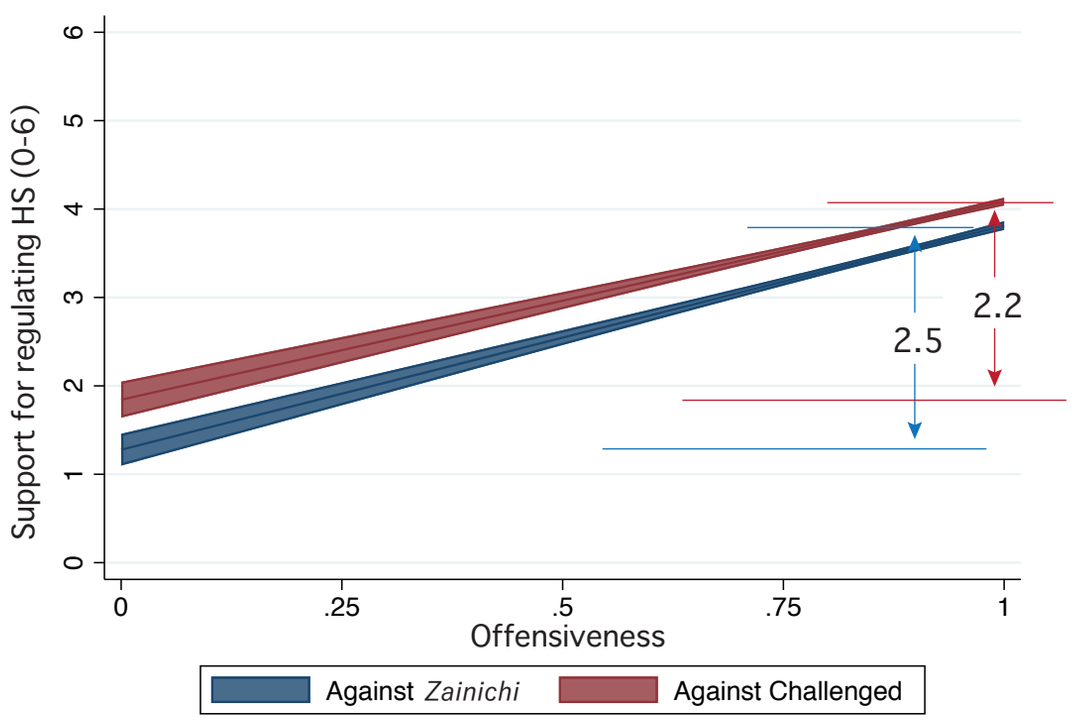


Figure 6: Support for Regulating HS -- Regression Analysis:
Effect of "Harm Dignity" by Target Groups

[***Zainichi*** N=4,077; Challenged N=4,149]

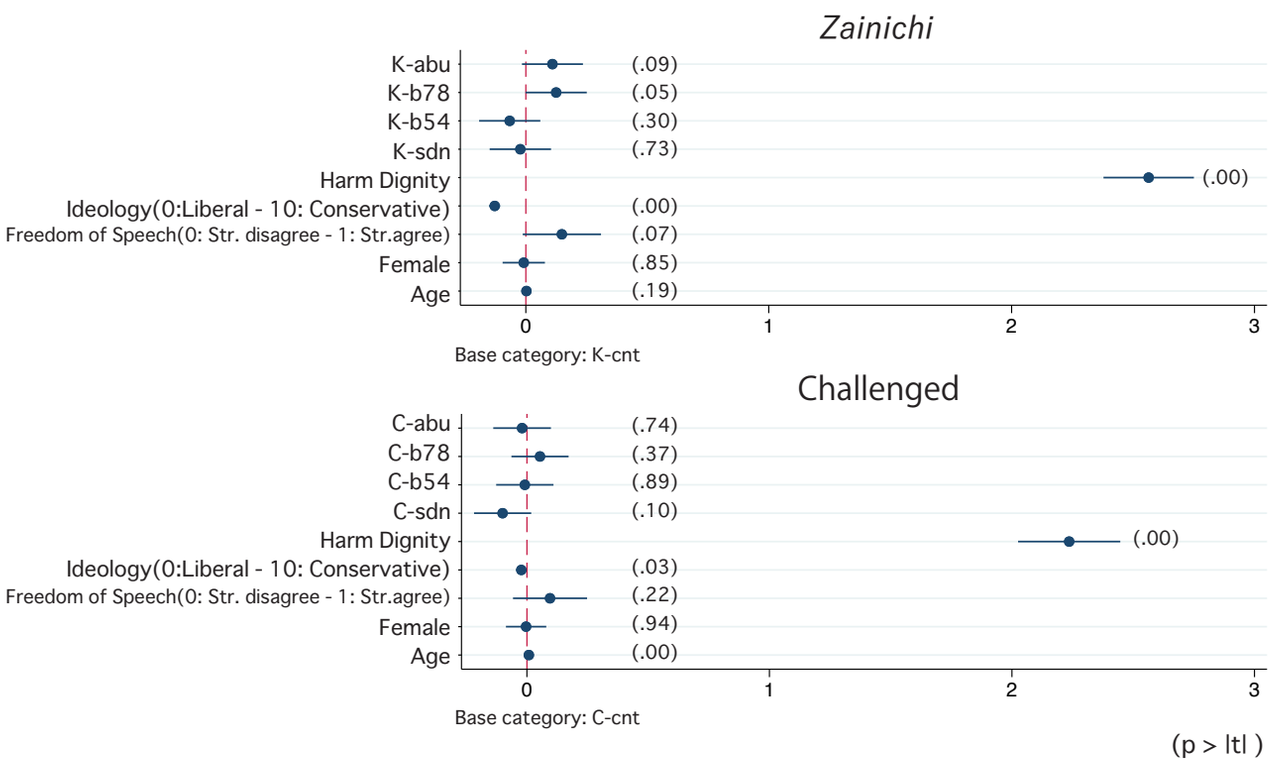


Figure 7: Support for Regulating HS with Bandwagon Effects
Post-estimation Simulation "Harm Dignity" by Target Groups

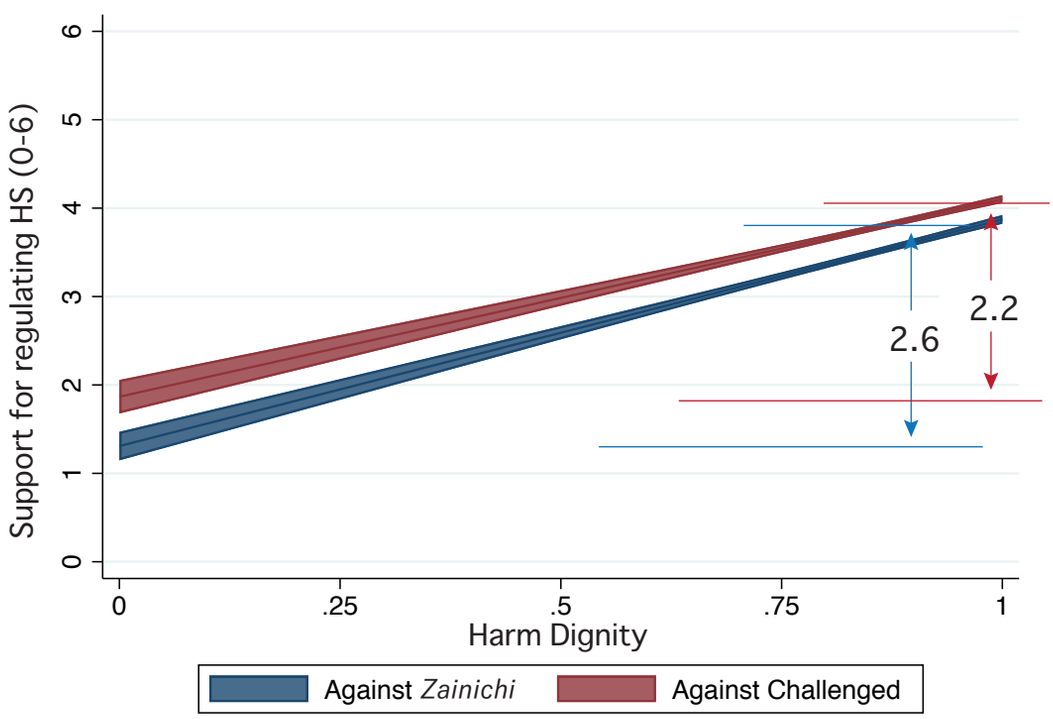


Figure 8: Support for Regulating HS -- Regression Analysis:
Effect of "Unfairness" by Target Groups

[***Zainichi*** N=4063; Challenged N=4,139]

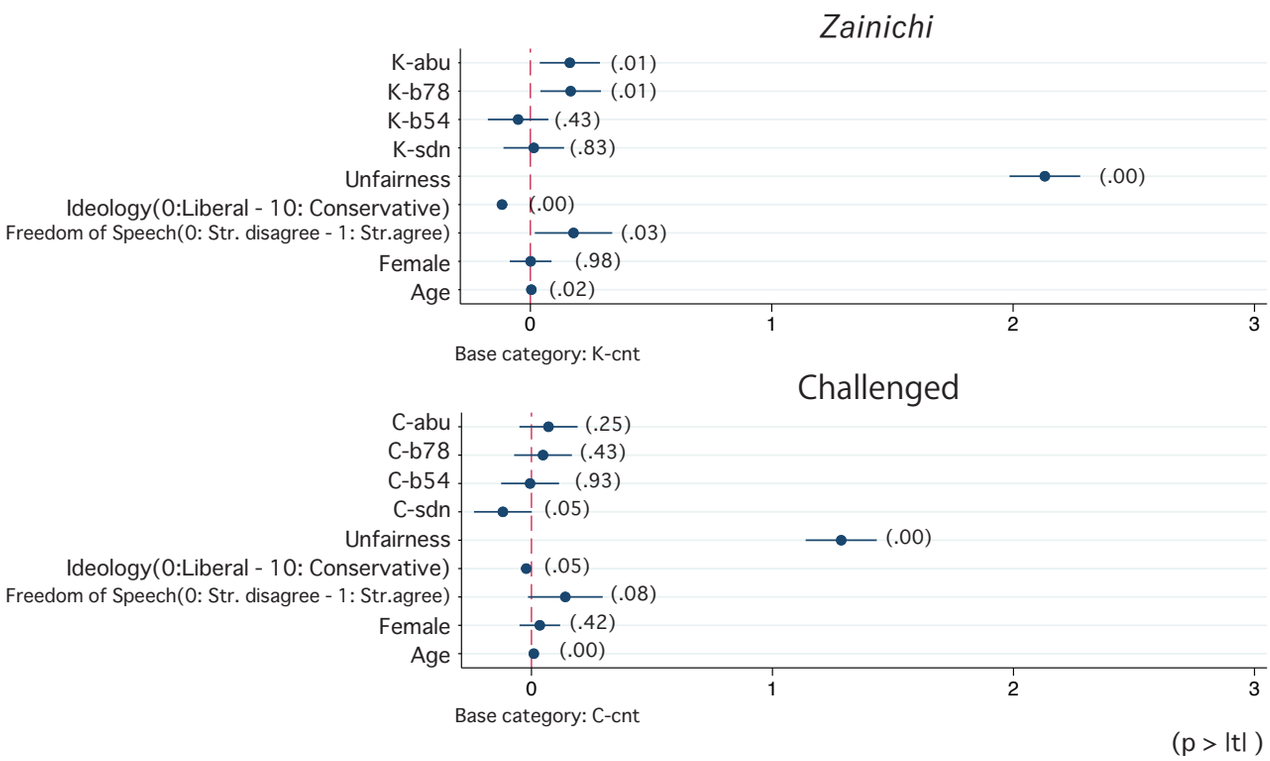


Figure 9: Support for Regulating HS with Bandwagon Effects
Post-estimation Simulation "Unfairness" by Target Groups

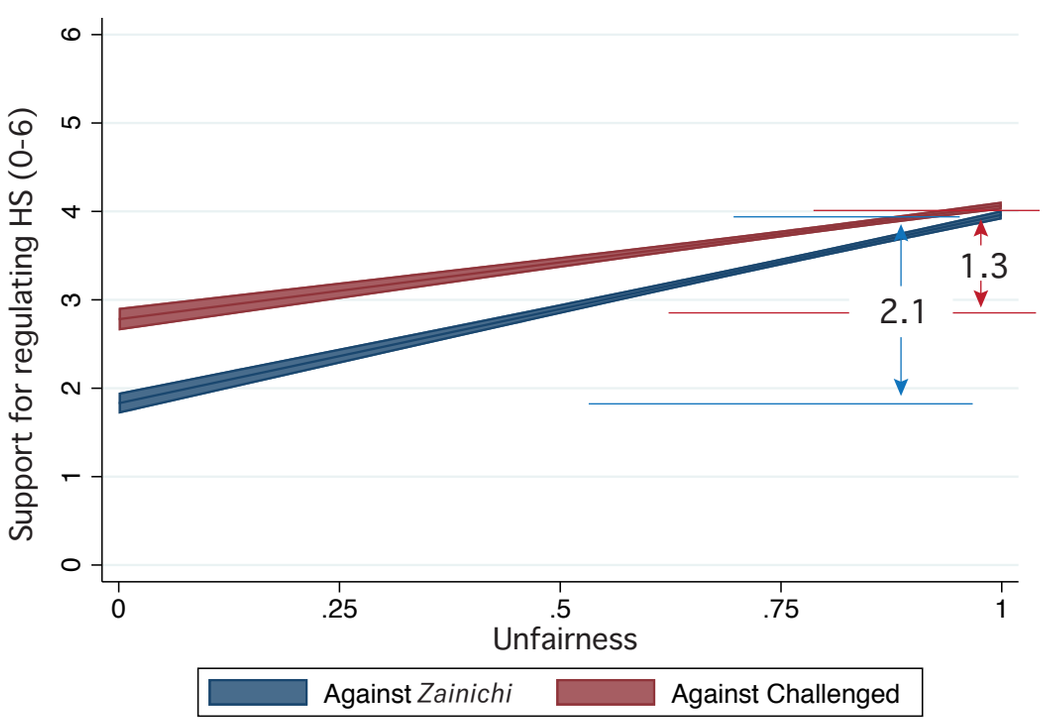


Figure 10: Support for Regulating HS -- Regression Analysis:
Effect of "Makes Society Worse" by Target Groups
[***Zainichi*** N=4,064; Challenged N=4,145]

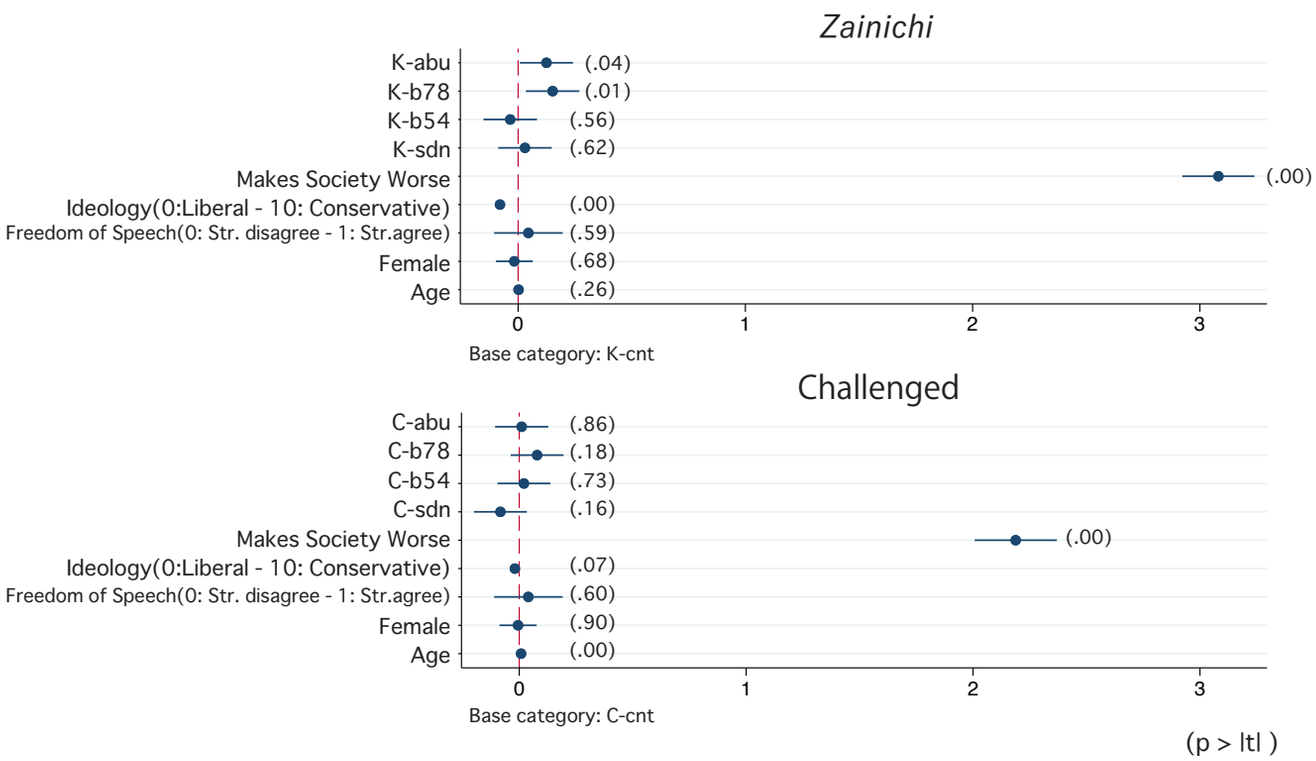
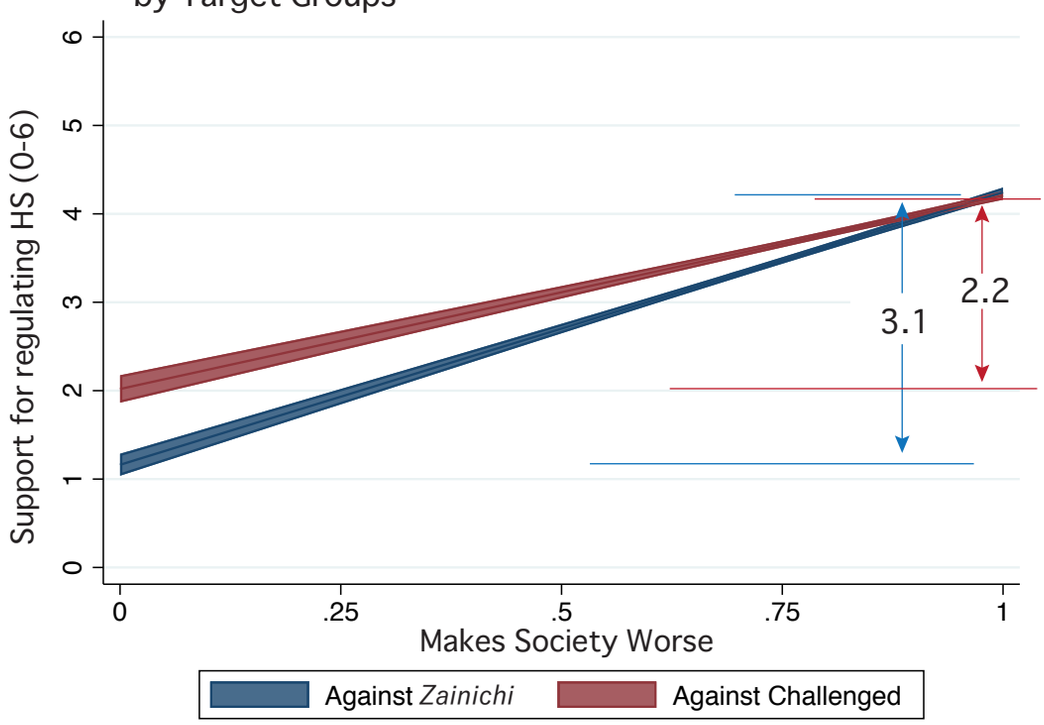


Figure 11: Support for Regulating HS with Bandwagon Effects
Post-estimation Simulation "Makes Society Worse"
by Target Groups



Appendix

Campaign Poster for “STOP! [sic] HATE SPEECH” Drive by the Japanese Ministry of Justice

http://www.moj.go.jp/content/001184410.pdf