Public Media Do Serve The State: A Field Experiment*

Shuhei Kitamura† Toshifumi Kuroda‡

August 23, 2019

Abstract: This paper analyzes the impact of public media on political attitudes. We have conducted a randomized field experiment in collaboration with the nation's public service broadcasting in which capacity for viewing its programs has been randomly increased. We find that the treatment has increased the evaluation on the government's foreign policies through an increase in the viewing time of its programs. To further study the mechanism, we develop a novel method to measure semantic similarity between TV programs and official statements made by domestic and foreign governments. We find that public media's information is semantically closer to the domestic government's statements than to the foreign government's compared to private media's information. In addition, we show that the positive evaluation on foreign policies is affected by individuals' being exposed to such information. In contrast to previous studies showing media slant and persuasion in private media outlets, this study adds a new empirical evidence that public media can persuade their viewers in support of the domestic government. We argue that the independence of public media may be as important as that of the Central Banks.

1 Introduction

Many countries have public media outlets including BBC in the UK, PBS in the US, and SVT and SR in Sweden. Their mission includes providing public services through media such as television and radio to citizens. In addition, a study conducted by the Reuters Institute at Oxford University shows that public media are most trusted in many countries including Denmark, Germany, Japan, and the UK (Newman et al., 2018). Also, PBS is most trusted by both Republicans and Democrats in the US (Pennycook and Rand, 2019). One of the reasons for this might be that public media are believed to have no particular political position compared to private media.

*This study is financially supported by JSPS grant numbers 16K03663 and 18K12768, and Japan Broadcasting Corporation. The views and opinions in this paper do not necessarily represent those of Japan Broadcasting Corporation. We thank Tomomichi Amano, Amy Catalinac, Matthew Gentzkow, Masayuki Kudamatsu, Vardges Levonyan, John List, Emi Nakamura, Torsten Persson, David Strömberg, Kensuke Teshima, Yasutora Watanabe, and Teppei Yamamoto for helpful comments and suggestions. We are also grateful for seminar participants at GRIPS, Keio, Osaka, Osaka City, and Rochester, and at the JEA Spring Meeting at Musashi University, the Kansai Labor Economics Workshop, the Kwansei Gakuin IO Workshop, the Kyoto Summer Workshop on Applied Economics, and the SIOE Conference in Stockholm. All errors are our own.

†Osaka University. Email: kitamura@osipp.osaka-u.ac.jp.

‡Tokyo Keizai University. Email: kuroda@tku.ac.jp.

A growing literature on the media slant shows the slant in private media (Gentzkow and Shapiro, 2010; Gentzkow et al., 2015) as well as controlled media (Enikolopov et al., 2011; Durante and Knight, 2012; Chen and Yang, 2019; Qin et al., 2018). Moreover, the persuasive effect of slanted private media has been documented in many studies (DellaVigna and Kaplan, 2007; Gerber et al., 2009; Schroeder and Stone, 2015; Martin and Yurukoglu, 2017; Durante et al., 2019). However, the study on public media's information and influence has been very limited. Is public media's information different from that of private media? If so, how and why? Is it also persuasive?

Given that public media co-exist with private media in many countries, studying public media's information and influence should complement earlier studies on private media just described. For example, theories indicate that the equilibrium information of private media might change, depending on the existence of public media in the market. If public media's information is less slanted than private media's, the existence of public media might reduce slanted information in equilibrium through either cross-checking (Mullainathan and Shleifer, 2005), or by providing a honest report (Gentzkow and Shapiro, 2006).[1] In contrast, if it is also slanted, public media might influence the public opinion more easily as they are the most trusted media in many countries (Newman et al., 2018). Therefore, it becomes crucial to check the neutrality of public media's information.

To the authors' best knowledge, ours is the first to show experimental evidence on the persuasive effect of public media on their viewers. In collaboration with the nation's public service broadcasting (henceforce, PSB) in Japan, we conducted a large-scale randomized experiment in the greater Tokyo area in 2016. The PSB is a public media outlet, which is not owned by the state and its revenue mostly comes from "receiving fees", which are paid by all households and enterprises that have TV equipment.[2]

Our experiment is designed to randomly increase capacity for viewing the PSB's TV programs. Using the experiment, we estimate the causal effect on evaluations on the government's capability for implementing policies and its measures against certain foreign affairs. We find that the treatment has increased the evaluation on the government's foreign policies by viewing the PSB's programs longer. However, we do not find any effect on the general evaluation on the government's capability.

To further study the mechanism, we use a novel method in which we apply unsupervised machine learning to measure semantic similarity between the contents of TV programs and that of official statements made by domestic and foreign governments. Specifically, we get our machine to learn semantic similarity between texts in these documents using a large corpus that is easily accessible for researchers.

Our method has two advantages. First, it does not require for researchers to select words of interest by themselves. Previous studies have often used pre-determined word categories made by scholars (Tetlock, 2007; Qin et al., 2017, 2018). With these methods, it is often difficult for researchers to pick up relevant words (e.g., finding party-specific words for political candidates). Moreover, the process might be time consuming. In contrast, computing semantic similarity between words in our method does not involve the process of selecting words *ex ante* and the machine learns semantic similarity just by using a corpus

[1] Strömberg (2015) also shows that competition leads to the under-provision of information compared to the social optimum, which can be alleviated by public media or by a Pigouvian subsidy.

[2] This is similar to BBC in the UK.

(i.e., a list of plain texts such as Google Books Corpora).

That specifying words is not needed in the process means that, as long as a relevant corpus is available, we can measure semantic similarity between *any* documents in theory such as official documents made by political actors such as bureaucrats and international organizations for which it is difficult to find specific words that represent them. For example, it might be very difficult to find words that represent an international organization such as the United Nations (henceforth, UN). Without specifying words that represent the UN in a document, our method can still measure how much a specific (possibly non-UN) document is semantically closer to an UN document.

Second, compared with similar methodologies employed in previous studies such as counting the frequency of words in texts (bag-of-words/bag-of-n-grams), our algorithm takes into account the ordering and semantics of words, as well as the contextual difference across documents.[3] Regarding the latter point, roughly speaking, our algorithm employs a "fixed-effect" procedure, in which the estimated parameter captures the contextual specificity of a document, when computing the probability of observing a word in a context.

To the best of our knowledge, this is the first study applying a method of measuring semantic similarity to the context of social sciences. Since the method can be used to measure distance between any texts as just described, we believe that it has many applications in economics and other subjects in social sciences.

Using our measure of similarity, we find that the PSB's programs are closer to domestic government's statements than to foreign government's compared to private media's programs. Moreover, the instrumental variables method confirms that the individuals tend to positively evaluate the domestic government's measures against foreign affairs by being exposed more to such information during the experiment.

This paper relates to a large body of the literature on media and information in economics and political science. First, it relates to the literature on the source of the media slant. The slant can be driven by either supply factors (e.g., owners, advertisers, journalists), or demand factors (consumers). Besley and Prat (2006) show that the incumbent is more likely to "buy" unfavorable information for her from the media outlet in order not to disclose information to voters if it is less costly relative to rents from office holding. Empirical studies find that controlled media, either by a state, a party, or a politician, are slanted in China (Qin et al., 2018; Chen and Yang, 2019), Berlusconi's Italy (Durante and Knight, 2012), and Russia (Enikolopov et al., 2011).

In contrast, Gentzkow and Shapiro (2010) show that the media slant in newspapers is driven by consumer preferences rather than owners in the US. Gentzkow et al. (2015) also find little evidence that the ruling party influences the partisan composition of the press. Our findings indicate that the information of public media, in contrast to that of private media, may be more susceptible to media capture as they can collect receiving fees without worrying any reduction in consumer demand.

In addition, our paper complements previous studies on persuasion by private media outlets, which find the impact of the slanted TV channel on the vote share (DellaVigna and Kaplan, 2007; Martin and Yurukoglu, 2017; Durante et al., 2019) and political knowledge (Schroeder and Stone, 2015). In contrast, a field experiment shows no impact of slanted

[3] Peterson and Spirling (2018) show the performance of different supervised machine learning algorithms for measuring polarization.

newspapers on the vote share and political knowledge and opinions, although it shows some evidence on support for Democrats and turnout (Gerber et al., 2009).

The literature on the impact of new information technology on political outcomes is also related because our experiment involves online streaming. Previous studies show that voter turnout decreases due to the diffusion of Cable TV (Gentzkow, 2006) and the Internet (Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2018). Using data from the US, Prior (2005) finds that new information technology has widened the gap of knowledge and turnout between individuals who prefer news and those who prefer entertainment. In contrast, Boxell et al. (2017) find that the greater use of the Internet is not associated with political polarization in the US.

Finally, our study relates to the literature on the analysis of political language. Previous studies use "word scores" (Laver et al., 2003), think-tank citations (Groseclose and Milyo, 2005), and party phrases (Gentzkow and Shapiro, 2006; Jensen et al., 2012; Martin and Yurukoglu, 2017; Gentzkow et al., 2019). As mentioned above, our method has several advantages such as measuring semantic similarity between documents without selecting words *ex ante*.

The rest of the paper is organized as follows. Section 2 explains background information and 3 describes our experiment and data. Section 4 shows the results from the experiment. Section 5 uses machine learning to further study the mechanism. Section 6 discusses the results and Section 7 concludes.

2 Background

2.1 Market Structure

The Japanese television market employs a dual structure, which consists of one nationwide public service broadcaster and several regional private broadcasters.

The PSB has two channels, 1 and 2, in any region. In addition, there are local private channels in each region, which broadcast the programs of five networks whose headquarters are all based in Tokyo and/or their own programs. No private broadcaster is allowed to have more than one channel in a region. In addition, the number of private broadcasters differs across regions. In the greater Tokyo area, for example, there are five major private channels; 4, 6, 8, 10, and 12. The entry and exit of private broadcasters in this area did not occur in the last 20 years.[4]

The PSB's market share varies from region to region as the number of private broadcasters also varies (from 2 to 6). In the greater Tokyo area, the market share of each channel, as measured by ratings, was Channel 1 (18.4%), 4 (22.5%), 6 (16.0%), 8 (15.5%), 10 (19.8%), and 12 (7.8%), respectively, in 2016.[5]

2.2 Source of Revenue and Ownership

The PSB is mostly (99% of annual income) funded by "receiving fees" from households and enterprises which have TV equipment regardless of whether they actually view the PSB's

[4] The last entry in the area occurred in 1995.

[5] Dentsu Media Innovation Lab (2018).

programs or not. It also receives government subsidy for providing international broadcasts and party election broadcasts. However, the total amount of the subsidy was only 0.5% of its annual income in 2016, when the House of Councillors election was held. Any revenue from advertisement is prohibited by the Broadcast Act (henceforth, the Act). The Act also rules that the PSB's annual budget must be approved by the Diet.

The twelve members of the Board of Governors, which decides the PSB's management policy and operations, are appointed by the Prime Minister and approved by the Diet.[6] The appointment sometimes comes to public attention. For example, the 2013 appointment gained media attention because some of the members were in too close relation to Prime Minister Abe (Nikkei, 2013). Historically, all the members have been Japanese nationals.

In contrast, the main source of revenue for private broadcasters is advertisement. It consisted of the 82.4% of their annual income in 2016.[7] Foreign nationals can buy the private broadcaster's share, although the Act regulates that foreign ownership with voting rights should not exceed 20%.[8]

2.3 Program Contents

Regarding the editing of broadcasting programs, the Act rules that the broadcaster "shall not harm public safety or good morals", "shall be politically fair", "shall not distort the facts", and "clarify the points at issue from as many angles as possible where there are conflicting opinions concerning an issue".[9] Although the same rule applies to all broadcasters, there are some differences in the program contents between the public and private broadcasters. Comparing news programs in the PSB and private broadcasting in the 1990s, Krauss (2000) finds that the frequency that journalists and newscasters express their personal opinions about a news in private broadcasting programs is twice as large as those in the PSB's programs. In addition, when the PSB newscasters make some comments on a particular news, they tend to describe the objective and obvious information rather than their personal opinions.

According to an online survey conducted by Newman et al. (2018), the PSB is the most trusted news outlet in Japan.[10] This style of the PSB's news programs may be a major reason of why it is most trusted by the public.

[6] They are typically business managers, academics, and celebrities. The term of appointment is 3 years, although it can be renewed.

[7] Japan Commercial Broadcasters Association (2017).

[8] If foreign ownership without voting rights is counted, the share of foreign shareholding can exceed 20% but in practice it does not deviate too much from this threshold.

[9] Ministry of Internal Affairs and Communications (2010)

[10] We also find that the PSB is more trusted than private broadcasting in our sample.

3 Experiment

3.1 Overview

A randomized controlled trial (henceforth, RCT) was conducted for the Internet users in the greater Tokyo area between November and December in 2016.[11] As the main treatment, we randomly provided free subscriptions to view the PSB's programs online.[12] In particular, the treated individuals were able to view the PSB's programs on web browsers and mobile phones. The services are comprised of live streaming and video on demand (henceforce, VOD). The treatment thus increased capacity for viewing the PSB's programs.

Although the PSB had not provided any live streaming service before the experiment began, it had provided a paid VOD service.[13] Some programs were not available during the experiment for reasons such as the copy right issue.[14] In addition, because of government regulation, only the programs that started at 7:00 am or after and ended by 11:00 pm were available for both services. In contrast, accessing the VOD service was possible at any time of the experiment.

We recruited 6,000 Internet users using a mail survey.[15] First, we made 40 blocks according to age (20s, 30s, 40s, 50s, and 60s), sex (male and female), and TV view frequency (almost everyday, at least once a week, at least once a month, and hardly ever or never). Then, we recruited individuals in the way that the percentage of each block in the sample becomes proportional to the population in each category in the greater Tokyo area.[16]

Recruitment was conducted through a major Internet-survey company, which is one of the largest Internet-survey companies in Japan in terms of the number of registered individuals. The company sent a recruiting email to its registered individuals to inform the outline of the surveys and rewards.[17] The email explained that the purpose was to survey the individual's daily viewing behavior of TV programs including private ones, daily media access, and evaluations on media, in particular TV. The email was also explicit about that participants

[11] Although some private broadcasters had provided either free or paid Internet streaming services, the PSB had not provided such a service until a revision of the Broadcast Act in 2014, which allowed the PSB to provide the Internet streaming service if, at least, it would not harm the market competition. In order to test that the streaming service meets the criterion, the PSB conducted large-scale randomized experiments in 2015 and 2016, for which this paper focuses on the 2016 experiment. Regarding the 2015 experiment, Kuroda et al. (2017) find no evidence that the streaming service harms the market competition. The size of the experiment was significantly larger in the 2016 experiment than the 2015 experiment. Moreover, video on demand (described below) was a new feature in the 2016 experiment.

[12] As the second experiment, we also sent program recommendations to randomly selected subgroups in the treatment group on every Fridays during the experiment. The Online Appendix provides analyses on this second treatment.

[13] A monthly subscription fee for the VOD service was 972 JPY (about 10 USD) per month or from 100 JPY (about 1 USD) per program.

[14] In total, about 16% of the total broadcasting time was not provided in these services.

[15] The number is decided to maximize statistical power given the restriction on budget and the rule imposed by the government.

[16] More specifically, the numbers are derived from the PSB's public-opinion poll conducted in the greater Tokyo area in July, 2016.

[17] We did not tell individuals the exact amount of the rewards. The company did not disclose the details of the rewards to us, but they were based on the company's incentive criterion, according to the company's reply. Most importantly, there was a bonus depending on the number of times the participants report the view record (see the data section).

would keep view record everyday for the purpose of the survey. To minimize selection bias, the availability of the free Internet streaming services was not informed at the time of the recruitment. Participants were told about it as a surprise just a few days prior to the start of the experiment. Finally, 5,000 individuals out of 6,000 were randomly assigned to the treatment group and the remainder were assigned to the control group.[18]

Figure 1 shows the timeline of the experiment. We started recruitment at the end of October in 2016. The baseline survey was conducted at the same time. We then randomized the sample. All participants have started keeping view record since November 14th. The free online services became available only for the treatment group on November 28th. Finally, the endline survey was conducted after the end of the experiment on December 18th.

3.2 Data

Our main data are each individual's view record. The participants received an email everyday from the survey company, in which they were asked to mark TV programs online that they viewed on that day. It included not only the PSB's programs, but also private broadcasters' programs. Although the data are self-reported, we try our best to diminish measurement errors they may cause. First, to reduce the reporting cost, the participants simply checkmarked the names of TV programs they viewed on the computer screen. Second, the participants were incentivized for reporting and were given an extra time before submitting the report.[19] Third, samples which seem unreliable are dropped as a robustness check. Finally, the instrumental variables method is employed. As explained above, the recording started two weeks before the start of the experiment.[20]

In the baseline survey, we asked individuals' socio-economic status, media usage, evaluations on media, and political orientation. In the endline survey, we asked evaluations on the streaming services and government's capability and policies, in addition to media usage and evaluations on media.

Moreover, we also assembled very unique data: TV scripts from all TV channels in the Tokyo region including private broadcasters. We collected the data using a TV digital recorder.[21]

Since 1,524 individuals (1,280 (25.6%) in the treatment group and 244 (24.4%) in the control group) did not complete the endline survey, 23 people were not seemingly participating the experiment, and an individual's view record seems to be unreliable[22], our final

[18] We used the `randomize` command on Stata 14. We allocated more individuals than a half to the treatment group because of the second experiment, which is described and analyzed in the Online Appendix.

[19] The bonus was given to individuals at the end of the experiment and the amount depended on the times they reported view record. The survey company did not tell participants the exact formula for calculating the bonus in order to avoid any strategic behavior. The participants were given three days to complete the report.

[20] For the treated individuals only, the access log of the PSB's Internet streaming service is also available. We also conduct an analysis using the data in the Online Appendix.

[21] The machine provides API to extract program information including the titles, descriptions, and scripts. We used Python to obtain information from the recorder and saved them in CSV format. Because scripts for live programs are typically typed by an operator or are generated by machine recognition, there can be a few second delay from actual voice.

[22] We have dropped an individual who has viewed almost all TV programs, according to the record.

sample became 4,453 (74.2%).[23]

We conduct balance checks between the treatment and control groups in the final sample using variables from the baseline survey, and report the results in the Online Appendix. Most variables are balanced between the two groups from which we conclude that the randomization was successful. Since there are a few variables that are not balanced[24], we include them as the baseline control in the following analysis.

Table 1 shows the summary statistics of main variables.

4 Results

4.1 Average Treatment Effect

The first analysis examines whether the treatment increases the viewing time of the PSB's main channel, Public Channel 1. In particular, we run the following regression using Ordinary Least Squares (henceforth, OLS):

$$\Delta\text{Public\_1}_i = \alpha + \beta\text{Treatment}_i + \mathbf{X_i}\gamma + \varepsilon_i, \tag{1}$$

for individual i, where $\Delta\text{Public\_1}_i$ is the difference in the viewing time (in hours) of Public Channel 1 before and during the experiment, Treatment_i is the treatment dummy, $\mathbf{X_i}$ is a vector of control variables, and ε_i is the error term, which we assume is independently distributed across individuals. Our main interest is the estimate on β. By using the difference in the viewing time, we remove any systematic difference in the usual viewing pattern across individuals that we cannot perfectly control.

Figure 2 shows the average values of the dependent variable for the treatment and control groups. It shows that the treatment group viewed Public Channel 1 approximately 1 hour longer than the control group during the experiment.

Table 2 shows the average treatment effect on the viewing time of Public Channel 1. The first column does not include any control variable described in the data section, while the next two columns include them. The third column drops individuals who did not occasionally report the viewing record for certain programs, although they were likely to view them, according to the access log. Overall, the treatment increased the viewing time of Public Channel 1 by about one hour, corresponding to Figure 2. In the following analysis, we use the smallest sample to minimize measurement errors.

4.2 Effect by Week and on Other TV Channels

How is the effect during the experiment decomposed by week? Table 3 shows the effect by week.[25] The effect appears in all three weeks. Moreover, the size of effects does not vary much across weeks, although the effect in Week 2 is slightly larger than the effect on the

[23] We also discuss the possibility of attrition bias in the Online Appendix.

[24] They are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, and trust and accuracy measures.

[25] Since there is no comparable weekly viewing time before the experiment, the viewing time before the experiment is now included as an additional control variable in the regressions.

other weeks. This shows that the treatment effect does not simply capture the effect on special daily programs.

Next, to understand the overall impact of our treatment, we investigate the effect on other TV channels. As already mentioned, there are six other channels: 2, 4, 6, 8, 10, and 12. We are particularly interested in knowing whether the increase in the viewing time of Public Channel 1 is explained either by a decrease in the viewing time of another channel, or by an increase in the total viewing time of TV programs.

Columns (1) through (7) in the top panel of Table 4 shows the average treatment effect by channel. The dependent variables use the difference in the viewing time as in equation (1), but for each channel. Column (1) replicates the column (3) in Table 2 for comparison. In column (2), we find that the treatment did not increase the viewing time of Public Channel 2, which is another channel broadcast by the PSB. Public Channel 2 is slightly different from other channels in the sense that it focuses on educational programs. We also find that it was the least popular channel in our data.[26] These factors might explain why the treatment did not increase the viewing time of Public Channel 2.

Turning to private channels, columns (3) through (7) show the treatment effect for each private broadcaster. We find that the treatment effect for Private Channels 4 and 12 is negative, although the latter effect is relatively weak. By contrast, the last column of the top panel in Table 4 shows that the treatment did not affect the total viewing time of TV programs. Taken together, these results indicate that the increase in the viewing time of Public Channel 1 is to some extent explained by the change in the viewing time of private channels, holding the total viewing time constant. Using difference in shares instead of difference in hours for the dependent variables, the bottom panel of Table 4 shows a consistent picture; the treatment also increased the share of Public Channel 1 and decreases the share of Private Channel 4. The effect on Private Channel 12 now comes out as insignificant.

4.3 Effect on Political Evaluations

Next, we turn to investigating the effect of viewing Public Channel 1 on political evaluations. Two major political events occurred at the time of the experiment, both of which are related to foreign policies. The first major event was President Vladimir Putin's visit to Japan between December 15th and 16th for attending the summit meeting. One of the main topics of the meeting was the Northern Territories dispute, a territorial dispute between Japan and Russia over four northern islands (Etorofu, Kunashiri, Shikotan, and Habomai). The main interest of the Japanese public at the time was whether there would be any progress on that issue during the meeting (The Economist, 2016; Chugrov and Streltsov, 2017). Another major event was that Prime Minister Shinzō Abe (henceforth, PM) announced that he would visit Pearl Harbor with President Barack Obama (at the time) to console the souls of war victims. Although the event actually occurred after the experiment (December 26-28th), the first public announcement was made on December 5th, i.e., during the experiment.

Figure 3 shows the popularity of each topic as measured by the frequency of Google search in Japan according to Google Trends. It shows that "真珠湾" (Pearl Harbor) became more popular when there was the first public announcement by PM, and that "北

[26] The average viewing time of Public Channel 2 during the experiment was only about 1 hour, while that of Public Channel 1 was about 11 hours, according to Table 1.

方領土" (Northern Territories) became more popular when there was the summit meeting. For comparison, searching results for "交通事故" (traffic accident) is also presented, which does not show any systematic pattern.

Since these two were the major political events at the time of the experiment, we intentionally included survey questions related to these events in the endline survey. Specifically, a survey question asks: "How do you evaluate the government's measures against the Northern Territories dispute?", and another question asks: "How do you evaluate Prime Minister Abe's visiting Pearl Harbor with President Obama to console the souls of war victims?" Individuals rated each question by specifying a number from 0 (the lowest rating) to 10 (the highest rating). In addition, we also asked individuals to evaluate the capability of the Japanese government for implementing policies. In particular, a survey question asks: "Do you evaluate that the current government has capability to implement policies?" There are four possible answers from agree to disagree.

To check the effect of viewing Public Channel 1 on these outcomes, we run the following regression using OLS:

$$\text{Evaluation}_i = \phi + \psi \Delta \text{Public\_1}_i + \mathbf{X_i} \xi + v_i, \tag{2}$$

for individual i, where Evaluation_i is a political evaluation just as described, $\Delta\text{Public\_1}_i$ is the difference in the viewing time of Public Channel 1 as in equation (1), $\mathbf{X_i}$ is a vector of baseline control variables, and v_i is the error term, which we assume is independently distributed across individuals. We are interested in the estimate on ψ.

By simply regressing the model (2), the estimated ψ may be biased because $\Delta\text{Public\_1}_i$ is likely to be endogenous. The bias can work in either direction. On the one hand, if individuals who view the channel longer tend to positively evaluate the government, the effect is likely to be overestimated. On the other hand, if individuals who view the channel longer tend to be skeptical about the government, the effect is likely to be underestimated. To isolate the causal effect, we use the treatment variable as the instrument for $\Delta\text{Public\_1}_i$ in the instrumental variables (henceforth, IV) framework.

Columns (1) through (3) in Table 5 show the results. The top panel shows the OLS estimates of equation (2). The results are not statistically significant for all dependent variables. Next, the middle panel reports the IV (2SLS) estimates, in which the treatment variable is used as the instrument. Although viewing Public Channel 1 longer affects neither the evaluation on the government's capability nor the evaluation on PM's visiting Pearl Harbor (columns (1) and (2), respectively), it has a statistically significant effect on the evaluation on the government's measures against the Northern Territories dispute (column (3)). That is, the treatment group is more likely to positively evaluate the government's effort against the territorial dispute than the control group. The first-stage F-statistic is slightly larger than the conventional level for the case of a single instrument. We also conduct weak instrument robust inference, and report resulted 95% confidence intervals and p-value. Column (3) shows that the confidence intervals contain only positive values and the effect remains significant.

The estimate (0.243) implies that one standard deviation increase in the viewing time of Public Channel 1 (8.848) would increase the evaluation by 2.15, or approximately a half of the control mean. The effect is relatively large. Overall, the OLS estimates are underestimated suggesting the second channel of bias as described above.[27]

[27] Another interpretation might be that the average treatment effect is smaller than the local average

Finally, the bottom panel shows the estimates of reduced-form regressions. As expected, the Intention-to-Treat effect (henceforth, ITT) on the evaluation on the government's capability and PM's visiting Pearl Harbor is null, while the ITT on the evaluation on the Northern Territories dispute is statistically different from zero, and positive.

Figure 4 plots the probability distribution of these three outcomes for the treatment and control groups. To make these figures, dependent variables are regressed on baseline controls, and then the residuals are plotted. As expected, distributions do not differ between the two groups in terms of the evaluation on the government's capability and PM's visiting Pearl Harbor. In contrast, distributions for the Northern Territories dispute show that evaluations have been shifted to the right for the treatment group compared to the control group. Moreover, the difference is observed over a wide range of the domains.

4.3.1 Effect from Not Viewing Private TV Programs

As shown earlier, Table 4 indicates some substitution between Public Channel 1 and Private Channel 4. In this section, we analyze whether the effect on political evaluations also comes from *not* viewing Private Channel 4.

Columns (4) through (6) in Table 5 use the difference in the viewing time of Private Channel 4 instead of Public Channel 1 for the independent variable. The top panel shows OLS estimates, while the middle panel shows IV (2SLS) estimates for which the treatment dummy is used as the instrument.

According to the table, the OLS estimates for the evaluation on government's capability for implementing policies and on PM's visiting Pearl Harbor are positive and statistically significant, but these estimates are most likely biased due to endogeneity. In contrast, the IV estimates are significant only for the evaluation on the government's measures against the Northern Territories dispute, although the effect is relatively weak. Moreover, the sign of all IV estimates becomes negative. Because the F-statistic is smaller than the conventional level, the weak instrument robust inference is also conducted. The corresponding confidence interval in column (3) indicates that the true effect is likely to be smaller than zero. Since we know from Table 4 that the treatment seems to have *decreased* the viewing time of Private Channel 4, the negative sign indicates that the positive evaluation on the Northern Territories issue may also come from by *not* viewing Nippon to some degree. However, in the next section, we show suggestive evidence that the result of positive evaluation is most likely due to viewing Public Channel 1 more rather than viewing Private Channel 4 less.

In the next section, we use scripts data to investigate information in the TV programs to which individuals were exposed during the experiment to understand why the treatment had a significant positive effect on the evaluation on certain foreign policies.

treatment effect. However, given that the independent variable is most likely endogenous, this is less likely to be the main reason.

5 Mechanism

5.1 Closeness to Official Statements

Given that our main outcomes are about foreign policies, we hypothesize that the treated individuals might be exposed more to information closer to what the domestic government provides than to what the foreign government does.

To test the hypothesis, we apply an unsupervised machine learning algorithm to compute semantic similarity between TV programs and official statements made by domestic (Japan) and foreign governments (USA and Russia). Then, we calculate the sum of the similarity measure, weighted by the duration of each program, for each individual. In other words, the variable captures the average exposure to certain information.

Formally, for individual i and official statement $j \in \{\text{Japan}, \text{ USA}, \text{ Russia}\}$, we compute

$$\text{AvgExposure}_{ij} := \sum_{p} \{1[\text{View}_{ip}] \times \text{Duration}_p \times \text{Closeness}_{pj}\}, \tag{3}$$

where $1[\text{View}_{ip}]$ is an indicator variable, which takes the value one if i viewed program p and zero otherwise, Duration_p denotes the duration of program p measured in hours, and Closeness_{pj} is the similarity between program p and the official statement made by government j, which takes the value between -1 (very far) and 1 (very close). The larger the value of Closeness_{pj}, the more similar are program p and government j's official statements. We use machine learning to construct this closeness variable.[28]

Next, we take difference in these measures between the domestic government and foreign government $k \in \{\text{USA}, \text{ Russia}\}$

$$\Delta\text{AvgExposure}_{ik} := \text{AvgExposure}_{i\text{Japan}} - \text{AvgExposure}_{ik}. \tag{4}$$

Intuitively, $\Delta\text{AvgExposure}_{ik}$ is made to capture average exposure to the official statement made by the domestic government, relative to average exposure to the official statement made by the foreign government. The variable takes a larger positive value if individuals are exposed more to information that is closer to what the domestic government provides than to what the foreign government does.

5.2 Construction of the Closeness Variable

To construct a variable Closeness_{pj} in (3), we apply a machine learning algorithm that learns fixed-length feature representations from pieces of texts (Le and Mikolov, 2014). We train a model such that it predicts the word based on the current context (e.g., predict the forth word from "the", "cat", and "sat").[29] Once the process is done, we compute similarity between each official statement and each TV program that was broadcast during the experiment, using semantic similarity between words contained in these documents. Using this definition, therefore, two documents are more similar if they contain more words that are semantically more similar.

[28] The procedure is described in the next section.

[29] This is also called the "Paragraph Vector with Distributed Memory (PV-DM)" model.

Let $\mathscr{D} = \{D_1, ..., D_K\}$ be the set of documents (TV programs) used for training a machine, and $\mathbf{X}$ be the set of all words in the training corpus, with length n. Each document D_k contains a sequence of words x_1^k, x_2^k, ..., $x_{I_k}^k \in \mathbf{X}$. Let $x_i^k \in \mathbf{X}$ be the target word, and $\mathbf{c}_i^k \in \{0,1\}^{n\times 1}$ be the local context $x_{i-s}^k, ..., x_{i-1}^k, x_{i+1}^k, ..., x_{i+s}^k$ of the target, where s is the maximum window size, in document D_k. Each element c_{ij}^k of the vector takes the value one if word position j in $\mathbf{X}$ appears in the local context of x_i^k, and zero otherwise.

Next, denote $\mathbf{P} \in \mathbb{R}^{h\times n}$ be the projection matrix from input to a hidden space of length h and $\mathbf{Q} \in \mathbb{R}^{n\times h}$ be the projection matrix from the hidden space to output, where $(\mathbf{q}_x)^T$ be the row of $\mathbf{Q}$ for word x. Finally, let $\mathbf{D} \in \mathbb{R}^{h\times K}$ be a matrix where each column represents a memory of the document, where $\mathbf{d}^k$ represents the column of $\mathbf{D}$ for document D_k.

Using the softmax function, the probability of observing the target word x_i^k in document D_k given the local context and the memory is

$$P(x_i^k|\mathbf{c}_i^k, \mathbf{d}^k) = \frac{\exp(\mathbf{q}_{x_i^k}(\mathbf{P}\mathbf{c}_i^k + \mathbf{d}^k))}{\sum_{x\in\mathbf{X}} \exp(\mathbf{q}_x(\mathbf{P}\mathbf{c}_i^k + \mathbf{d}^k))}. \tag{5}$$

Then the two projection matrices are trained using stochastic gradient descent to minimize the loss function:

$$-\sum_{k=1}^{K}\sum_{i=1}^{I_k} \log P(x_i^k|\mathbf{c}_i^k, \mathbf{d}^k). \tag{6}$$

Given a trained $\mathbf{P}$, one can compute $\mathbf{d^k}$ by taking the average of columns in $\mathbf{P}$ of words included in document D_k. When the training is complete, words with similar semantics are mapped to a similar location in the vector space.[30]

As described in the introduction, a great advantage of the algorithm is that it takes into account the ordering and semantics of words, as well as contextual difference across documents. In our case, a document means a TV program. That is, we allow for a word in a program to be semantically different from the same word in another program. Moreover, the algorithm does not require *ex-ante* selection of words by researchers. Instead, the algorithm computes semantic similarity between words only using a corpus.

The data used for training our machine, i.e., a corpus, are the scripts of TV programs which were broadcast after the experiment ended until August 28th in 2018.[31] This choice of corpus is naturally made based on the feature of our study, but other studies may think of using other corpora such as Google Books Corpora, depending on the purpose of the study. We use Python's `gensim` package for all procedures.

Since Japanese texts, in contrast to western texts, are not separated as single words in the original data, we parse the scripts using a common Japanese parser called `MeCab`. We use only nouns, verbs, adjectives, and adverbs, so that our measures capture similarity between texts in a meaningful way.[32] Moreover, we apply stemming for all words. The total number of TV programs in training data is 195,016, which contain 613,497 distinct words.

[30] See e.g. Rong (2016).

[31] The reason for not including the experiment period in the training data is to avoid any direct connection between the texts that are used for training and the texts that we want to measure closeness. The reason for including only data until August 28th, 2018 is that the latest version of the dictionary that we use in the parsing process (NEologd) contains the data only until that date. Although the dictionary has been constantly updated, we avoid using the running version in order to make the replication possible.

[32] Excluded parts of speech are interjection, conjunction, particles, auxiliary verbs, prefix, and determiner.

For official statements, we use those made by the Japanese and American government when PM and President Obama visited Pearl Harbor together[33], and those made by the Japanese and Russian government when PM and President Putin met at the summit meeting in Japan.[34] For each case, the statements were about the same topic, and were spoken to the public almost at the same time in the same place. Thus, only the difference between official statements for each event is whether the statement was made by the domestic government or the foreign government.

Before starting, the algorithm requires researchers to decide a couple of parameters: window size (i.e., how many words are considered together in the moving window on texts, or s in the model) and vector size (i.e., how many dimensions are considered to capture the complex meanings of each word, or h in the model). To get the baseline numbers for these parameters, we compute the percentage of the closest programs to official statements that can be categorized as either the USA or Russia topic to see how much we are successful at predicting relevant programs.

In particular, we first make a list of TV programs and order them from the most closest to the least closest to a statement made in each event. We then select a few programs from the most closest programs and check the percentage of programs that can be categorized as either the USA or Russia topic.[35] Thus, a higher percentage means that we are more successful at predicting relevant programs for each topic. We choose the top 7, 35, and 70 closest programs, which roughly correspond to the 1%, 5%, and 10% of all TV programs in the data.

We start by using two window sizes, i.e., 5 and 15, for the moving window. Figure 5 plots a figure for window size 15, and Figure 6 shows a figure for window size 5. In both figures, the left figures use closeness to the Japanese government's statement (top) and the American government's statement (bottom), respectively. In contrast, the right figures use closeness to the Japanese government's statement (top) and the Russian government's statement (bottom), respectively.

These figures show that our machine with window size 15 is relatively worse at predicting programs for the USA topic, in particular when closeness to the American official statement is used. In addition, by looking at Figure 6, which uses window size 5, we find that the percentage roughly decreases as vector size increases for the top 7 programs if closeness to the Japanese official statement is used, while it increases for the top 7 programs if closeness to the American official statement is used, with the lowest hit when vector size is 100.

Based on these analyses, we have chosen window size 5 and vector size 200 for the main analysis. The Online Appendix provides the results with other parameter values.

[33] The official statements in Japanese are available at the Cabinet Office's website (http://www.kantei.go.jp/jp/97_abe/statement/2016/1227usa.html) and the US Embassy and Consulates in Japan's website (https://jp.usembassy.gov/ja/remarks-president-obama-and-prime-minister-abe-japan-pearl-harbor-ja/). The official statements in English are also available at the Cabinet Office's website (http://japan.kantei.go.jp/97_abe/statement/201612/1220678_11021.html) and the White House's website (https://obamawhitehouse.archives.gov/the-press-office/2016/12/28/remarks-president-obama-and-prime-minister-abe-japan-pearl-harbor) (last accessed: 2019.8.20).

[34] The official statements are only available in Japanese and are available at the Cabinet Office's website (https://www.kantei.go.jp/jp/97_abe/statement/2016/1216kaiken.html) (last accessed: 2019.8.20).

[35] See the Online Appendix for the procedure of categorizing these topics.

5.3 Results

Table 6 shows the difference in closenesses between programs and official statements without taking account the participants' viewing behavior in AvgExposure_{ij} in (3). The table is made to check the existence of difference in information from the supply side. To make the table, we first take, for each program, the difference in closesses to either government (domestic/foreign), then compare the averages between two types of broadcasters (Public Channel 1/private channels).

The top panel also takes into account the duration of the program, while the bottom panel does not. First of all, the table shows that all differences are statistically significant. Moreover, they are all positive, indicating that Public Channel 1's information is closer to the domestic government than to the foreign government compared to private broadcasters' information. Interestingly, the difference between Japan and Russia is somewhat larger than the difference between Japan and America.

Next, in Figure 7, we take into account participants' actual viewing behavior. Each point in the figure corresponds to the individual's average exposure to particular information, i.e., AvgExposure_{ij}. The exposure to the domestic government is plotted against the exposure to the foreign government. Red lines indicate 45-degree lines. Figure 7 shows the results for Public Channel 1.

Strikingly, the figure shows that data points tend to be located above the 45-degree lines, especially for individuals who are away from the origin, indicating that individuals are more likely to be exposed to information closer to the domestic government's information than to the foreign government's, by viewing Public Channel 1's programs. This result gets even starker when it is compared with Figure 8, which plots the same figures but for programs in private broadcasters (except for the top-left figures, which shows the results for Public Channel 1 for comparison). It shows that data points are more or less on the 45-degree lines. These results suggest that individuals have been affected by information somewhat unique to Public Channel 1 during the experiment.

To check whether or not the positive second-stage result on the political evaluations is explained by this informational difference, we next use the closeness measures in regressions.

The top panel of Table 7 shows the first-stage results in which the treatment dummy is used as the main independent variable and closeness between Public Channel 1 and official statements are used as the dependent variables. "Δ Japan-USA" means the difference in average closenesses between the Japanese and American governments, while "Δ Japan-Russia" is the difference in average closenesses between the Japanese and Russian governments.

Columns (1) and (4) include all TV programs broadcast by Public Channel 1; columns (2) and (5) use only the programs that are categorized as the topic related to PM; columns (3) and (6) use only the programs categorized as the Russia topic. Overall, the table shows that the treatment group is more likely to receive information closer to what the domestic government provides than to what the foreign government does. This is probably not a surprising result given the uniqueness of Public Channel 1's information just we saw.

Next, the top panel of Table 8 shows the OLS estimates in which the the evaluation on the Northern Territories dispute is used as dependent variables, i.e., the same dependent variables used in column (3) in Table 5, and the closeness measures are used as independent variables. The bottom panel of the same table shows the IV (2SLS) estimates in which the treatment variable is used as the instrument. The structure of the table is the same as

Table 7; the dependent variable uses either all programs or a subset of them focusing on a particular topic, depending on columns.

Although the effect is somewhat weak in some specifications, all IV estimates show a positive sign, indicating that by being exposed to information closer to the domestic government, individuals are more likely to positively evaluate the government's measures against the Northern Territories dispute. Since the F-statistic of the first-stage is smaller than the conventional level, we also conduct the weak instrument robust inference and report confidence intervals. Compared with the standard confidence intervals, these confidence intervals tend to be wider. However, they tend to only include positive values, indicating that the true effect is more likely to be positive than negative or null.

How about the magnitude? Using the estimate in column (1) (7.389), one standard deviation (0.681) increase in the independent variable increases the evaluation on the Northern Territories dispute by 5. Similarly, the estimate in column (4) (7.357) means that a one standard deviation increase in the treatment variable increases the evaluation by 5. Therefore, the effect size is considerably large.

Next, because in the previous section, we find that there seems to be some (albeit weak) second-stage effect through Private Channel 4, we also examine whether the effect on the political evaluation also comes from *not* being exposed to information closer to the foreign government. To check this possibility, we make closeness measures for Private Channel 4 using the same procedure described in the paper. The first-stage estimates of regressing these measures on the treatment dummy are shown in the bottom panel of Table 7. The table does not show clear evidence that treatment has induced individuals *not* being exposed to information closer to the foreign government. Compared with the top panel of the same table, the estimates are also relatively smaller. This is not surprising given that individuals do not seem to be exposed to slanted information through Private Channel 4, according to Figure 7.

Taken together, the positive evaluation on Northern Territories issue is more likely to be explained by the viewers being exposed *more* to information closer to the domestic government, rather than by them being exposed *less* to information closer to the foreign government.

Overall, we find that the treated individuals have positively evaluated foreign policies because they were exposed to information closer to what the domestic government provides during the experiment.

6 Discussion

Why are public media much closer to the domestic government than to the foreign government compared to private media? As already noted in the introduction, the previous literature shows that the media slant is driven by either (a) demand factors (Gentzkow and Shapiro (2010)), or (b) supply factors (Enikolopov et al. (2011); Durante and Knight (2012); Qin et al. (2018); Chen and Yang (2019)). Regarding (a), since receiving fees are collected regardless of viewers' political orientation, it is hard to believe that the consumers' political preferences are driving the uniqueness in public media's information. The same argument might be applied for public media in other countries such as Sweden and the UK in which similar fees or taxes are collected from almost all adults.

Turning to (b), since receiving fees are mandatory for anyone who has TV equipment, changing information in public media should not substantially affect their revenue. Furthermore, as mentioned in the background, their annual budget and board members need the approval of the Diet and they do not receive any revenue from commercial advertisement. These facts indicate that a most plausible explanation could be that public media are more susceptible to media capture by the government, if anything.

An alternative hypothesis could be that public media's information might be more slanted towards the ruling party than the opposition party compared to private media's information, and that such information might drive our results. We test this hypothesis in the Online Appendix, in which we do not find any result supporting this hypothesis. In other words, our results do not seem to be explained by political competition between political parties.

Finally, although we find that public media's information tends to be closer to the domestic government's information for both topics, we find the significant result only for the Northern Territories dispute in Table 5. A possible explanation might be that the summit meeting was just over when the endline survey was conducted, while the PM's visiting Pearl Harbor was not. The former event might have endorsed individuals' beliefs and hence have influenced their survey answers. An alternative explanation could be that our closeness measures are not perfect for eliciting information to which individuals were exposed during the experiment, although we think that they are the most intuitive measures we can think of.

7 Conclusion

In contrast to the previous literature on the media slant and its influence, which focuses on private media, this paper examines the slant and influence of public media in a democratic setting.

To estimate the causal effect of public media, we have conducted a randomized experiment in collaboration with the nation's public service broadcasting in which capacity for viewing it TV programs has been randomly increased. We find that the treatment has increased the evaluation on the government's foreign policies through an increase in the viewing time of the public TV's programs.

To understand the mechanism, we develop a novel method in which we apply machine learning to measure semantic similarity between the contents of TV programs and official statements made by domestic and foreign governments. The method has several advantages including no need for selecting words *ex ante* in order to measure semantic similarity between documents. This means that the method can be easily applied for other cases where representative words are difficult to found. Using the similarity measure, we find that treated individuals were exposed more to the information closer to what the domestic government provides than to what the foreign government does during the experiment, which explains the positive evaluation on the foreign policies.

A further analysis on the role of public media in other countries and settings may be needed. In addition, future studies should address the question on how humans change their beliefs and attitudes in detail by studying their neural response to information.

References

Besley, T. and A. Prat (2006). Handcuffs for the Grabbing Hand? Media Capture and Government Accountability. *American Economic Review 96*(3), 720–736.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences 114*(40), 10612–10617.

Campante, F. R., R. Durante, and F. Sobbrio (2018). Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation. *Journal of the European Economic Association 16*(4), 1094–1136.

Chen, Y. and D. Yang (2019). The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review 109*(6), 2294–2332.

Chugrov, S. V. and D. V. Streltsov (2017). Interdependence of russo-japanese relations and mutual images of japan and russia. *Japanese Journal of Political Science 18*(1), 22–40.

DellaVigna, S. and E. Kaplan (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Dentsu Media Innovation Lab (2018). *Jyouhou Media Hakusyo*. Tokyo: Diamond Inc.

Durante, R. and B. Knight (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association 10*(3), 451–481.

Durante, R., P. Pinotti, and A. Tesei (2019). The Political Legacy of Entertainment TV. *American Economic Review 109*(7), 2497–2530.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review 101*, 3253–3285.

Falck, O., R. Gold, and S. Heblich (2014). E-lections: Voting Behavior and the Internet. *American Economic Review 104*, 2238–2265.

Gavazza, A., M. Nardotto, and T. M. Valletti (2018). Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies. *Review of Economic Studies, forthcoming*.

Gentzkow, M. (2006). Television and Voter Turnout. *The Quarterly Journal of Economics 121*(3), 931–972.

Gentzkow, M., N. Petek, J. M. Shapiro, and M. Sinkinson (2015). Do Newspapers Serve The State? Incumbent Party Influence on The US Press, 1869–1928. *Journal of the European Economic Association 13*(1), 29–61.

Gentzkow, M. and J. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy 114*(2), 280–316.

Gentzkow, M. and J. M. Shapiro (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica 87*(4), 1307–1340.

Gerber, A. S., D. Karlan, and D. Bergan (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal: Applied Economics 1*(2), 35–52.

Groseclose, T. and J. Milyo (2005). A Measure of Media Bias. *The Quarterly Journal of Economics 120*(4), 1191–1237.

Japan Commercial Broadcasters Association (2017). *Nihon Minkan Housou Nenkan*. Tokyo: Kōken Publishing.

Jensen, J., E. Kaplan, S. Naidu, and L. Wilse-Samson (2012). Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech. *Brookings Papers on Economic Activity*, 1–81.

Krauss, E. S. (2000). *Broadcasting Politics in Japan: NHK and Television News*. Ithaca: Cornell University Press.

Kuroda, T., D. Kido, and Y. Ichikawa (2017). Does the usage of TV streaming on the Internet bring about new functions to the public service broadcaster?: Evidence from a randomized field experiment. *Mimeo*.

Laver, M., K. Benoit, and J. Garry (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review 97*(2), 311–331.

Le, Q. V. and T. Mikolov (2014). Distributed Representations of Sentences and Documents. *CoRR abs/1405.4*.

Martin, G. J. and A. Yurukoglu (2017). Bias in cable news: Persuasion and polarization. *American Economic Review 107*(9), 2565–2599.

Ministry of Internal Affairs and Communications (2010). The Broadcast Act. `http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Resources/laws/pdf/090204_5.pdf`. (Last accessed: 2019-01-28).

Mullainathan, S. and A. Shleifer (2005). The Market for News. *American Economic Review 95*(4), 1031–1053.

Newman, N., R. Fletcher, A. Kalogeropoulos, D. A. L. Levy, and R. K. Nielsen (2018). *Reuters Institute Digital News Report 2018*. Oxford: Reuters Institute for the Study of Journalism.

Nikkei (2013). NHK Keiei Iin Kokkai Doui. `https://www.nikkei.com/article/DGXNASDD080D5_Y3A101C1TJ0000/`. (Last accessed: 2019-01-26).

Olea, J. L. M. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Pennycook, G. and D. G. Rand (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences 116*(7), 2521–2526.

Peterson, A. and A. Spirling (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems. *Political Analysis 26*(1), 120–128.

Prior, M. (2005). News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout. *American Journal of Political Science 49*(3), 577–592.

Qin, B., D. Strömberg, and Y. Wu (2018). Media bias in China. *American Economic Review 108*(9), 2442–2476.

Qin, B., D. Strömberg, and Y. Wu (2017). Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda. *Journal of Economic Perspectives 31*(1), 117–140.

Rong, X. (2016). word2vec Parameter Learning Explained. *Mimeo*.

Schroeder, E. and D. F. Stone (2015). Fox News and political knowledge. *Journal of Public Economics 126*, 52–63.

Strömberg, D. (2015). Chapter 13 - Media Coverage and Political Accountability: Theory and Evidence. In S. P. Anderson, J. Waldfogel, and D. Strömberg (Eds.), *Handbook of Media Economics*, Volume 1 of *Handbook of Media Economics*, pp. 595–622. North-Holland.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance 62*(3), 1139–1168.

The Economist (2016). Two men in a tub. `https://www.economist.com/asia/2016/12/10/japans-prime-minister-plans-a-steamy-tete-a-tete-with-russias-president`. (Last accessed: 2019-01-01).

Tables and Figures

Table 1: Summary statistics.

	Mean	Std.	Min	Max	Obs.
Total view during the experiment (hour)					
Public_1	11.03	19.87	0.00	174.55	4452
Public_2	1.03	3.86	0.00	74.22	4452
Private_4	25.23	33.38	0.00	413.88	4452
Private_6	17.03	23.02	0.00	309.90	4452
Private_8	16.77	25.94	0.00	230.25	4452
Private_10	22.26	28.68	0.00	362.97	4452
Private_12	6.01	9.34	0.00	130.12	4452
Difference in total view					
Δ Public_1	0.94	8.69	−44.97	116.07	4452
Δ Public_2	0.24	2.14	−17.37	74.22	4452
Δ Private_4	8.16	15.13	−62.15	134.75	4452
Δ Private_6	5.18	11.27	−66.05	123.45	4452
Δ Private_8	5.79	11.76	−56.00	113.37	4452
Δ Private_10	7.45	13.16	−70.98	135.70	4452
Δ Private_12	1.24	4.97	−34.45	41.32	4452
Closeness					
Δ Japan-USA	0.37	0.66	−0.28	6.43	4452
Δ Japan-Russia	0.41	0.75	−0.34	6.26	4452
Treatment					
Treatment	0.83	0.37	0.00	1.00	4452
Ad.	0.67	0.47	0.00	1.00	4452
Ad. + info.	0.33	0.47	0.00	1.00	4452
Political evaluation					
Gen. eval.	2.33	0.88	1.00	4.00	4452
Pearl Harbor	6.85	2.54	0.00	10.00	4452
North. Terr.	4.51	2.55	0.00	10.00	4452

Table 2: The average treatment effect on viewing Public Channel 1.

	Dependent variables: Δ Public_1		
	(1)	(2)	(3)
Treatment	0.815	0.887	1.082
	$(0.329)^{**}$	$(0.331)^{***}$	$(0.336)^{***}$
Controls	No	Yes	Yes
Mean of control	0.27	0.27	0.27
R^2	0.00	0.01	0.01
Observations	4452	4452	4156

Notes: Robust standard errors are in parentheses. * indicates p <0.10, ** p <0.05, and *** p <0.01. Dependent variables are the difference in the viewing time of Public Channel 1 before and during the experiment (in hours). "Treatment" is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, and trust and accuracy measures. Column (3) drops individuals who did not occasionally report the viewing records of certain programs, although they were likely to seeing them according to the access logs during the experiment.

Table 3: The average treatment effect on viewing Public Channel 1, by week.

	Dependent variables: Public_1		
	(1) Week 1	(2) Week 2	(3) Week 3
Treatment	0.300	0.420	0.321
	$(0.125)^{**}$	$(0.125)^{***}$	$(0.125)^{**}$
Controls	Yes	Yes	Yes
Mean of control	3.72	3.36	3.47
R^2	0.80	0.78	0.77
Observations	4156	4156	4156

Notes: Robust standard errors are in parentheses. * indicates p <0.10, ** p <0.05, and *** p <0.01. Dependent variables are the weekly viewing time of Public Channel 1 during the experiment (in hours). Week 1: November 28th-December 4th, Week 2: December 5th-December 11th, and Week 3: December 12th-December 18th. "Treatment" is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, trust and accuracy measures, and the viewing time of Public Channel 1 before the experiment.

Table 4: The average treatment effect on viewing TV programs, by channel.

	Dependent variables: Δ TV view							
	(1) Public_1	(2) Public_2	(3) Private_4	(4) Private_6	(5) Private_8	(6) Private_10	(7) Private_12	(8) All
Level								
Treatment	1.082	−0.028	−1.538	−0.513	−0.237	−0.409	−0.355	−1.999
	(0.336)***	(0.077)	(0.624)**	(0.463)	(0.462)	(0.548)	(0.209)*	(1.473)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	0.27	0.24	9.28	5.44	5.81	7.74	1.51	30.28
Mean of control (before)	10.28	0.71	17.96	11.16	10.24	15.06	5.04	70.44
Mean of control (during)	10.54	0.95	27.23	16.60	16.05	22.80	22.80	100.72
R^2	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.01
Observations	4156	4156	4156	4156	4156	4156	4156	4156
Share								
Treatment	0.014	−0.001	−0.010	−0.006	−0.001	0.003	0.001	
	(0.005)***	(0.001)	(0.005)**	(0.004)	(0.004)	(0.005)	(0.003)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Mean of control	-0.04	0.00	0.02	0.01	0.01	0.01	-0.01	
Mean of control (before)	0.18	0.01	0.24	0.15	0.14	0.21	0.08	
Mean of control (during)	0.14	0.01	0.26	0.16	0.15	0.22	0.07	
R^2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	
Observations	3934	3934	3934	3934	3934	3934	3934	

Notes: Robust standard errors are in parentheses. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Dependent variables are the difference in the viewing time of TV programs before and during the experiment for each channel, except for column (8) for which the dependent variable uses the total viewing time of all channels (in hours). "Treatment" is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, and trust and accuracy measures.

Table 5: Effect of viewing TV programs on political evaluations.

	Dependent variables: Evaluation					
	Public_1			Private_4		
	(1) Gen. eval.	(2) Pearl Harbor	(3) North. Terr.	(4) Gen. eval.	(5) Pearl Harbor	(6) North. Terr.
OLS Estimates						
Δ TV view	0.002	0.006	−0.002	0.002	0.011	−0.000
	(0.001)	(0.004)	(0.004)	(0.001)**	(0.002)***	(0.002)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	2.34	6.90	4.41	2.34	6.90	4.41
R^2	0.23	0.11	0.11	0.23	0.12	0.11
Observations	4156	4156	4156	4156	4156	4156
IV Estimates						
Δ TV view	0.042	0.044	0.243	−0.030	−0.031	−0.171
	(0.031)	(0.090)	(0.115)**	(0.023)	(0.065)	(0.092)*
Controls	Yes	Yes	Yes			
Mean of control	2.34	6.90	4.41	Yes	Yes	Yes
F-stat.	10.38	10.38	10.38	6.07	6.07	6.07
Effect. F-stat.	10.38	10.38	10.38	6.07	6.07	6.07
p-value (AR)	0.14	0.62	0.01	0.14	0.62	0.01
CI (AR)	[-.013, .144]	[-.153, .284]	[.074, .696]	[-.121, .010]	[-.285, .121]	[-.531,-.051]
CI (Wald)	[-.019, .103]	[-.133, .221]	[.017, .469]	[-.075, .016]	[-.158, .096]	[-.351, .009]
Observations	4156	4156	4156	4156	4156	4156
Reduced-form						
Treatment	0.046	0.048	0.263			
	(0.031)	(0.097)	(0.095)***			
Controls	Yes	Yes	Yes			
Mean of control	2.34	6.90	4.41			
R^2	0.23	0.11	0.12			
Observations	4156	4156	4156			

Notes: Robust standard errors are in parentheses. * indicates p <0.10, ** p <0.05, and *** p <0.01. Effective F-stat. is the effective F-statistic of Olea and Pflueger (2013). We use Stata's `weakivtest` for computing the test statistic. Weak instrument robust inference (Anderson-Rubin (AR) test) is conducted and the corresponding p-value and confidence interval are reported in rows labeled as "p-value (AR)" and "CI (AR)", respectively. We use Stata's command `rivtest` for computing the test statistic. Dependent variables are: (a) the evaluation on government's capability ("Do you evaluate that the current government has capability to implement policies?" (4: Yes, 3: More of less, 2: Not quite, 1: No)) (columns (1) and (4)), (b) the evaluation on Prime Minister Abe's visiting Pearl Harbor with President Obama ("How do you evaluate Prime Minister Abe's visiting Pearl Harbor with President Obama to console the souls of war victims?" (0 to 10 points)) (column (2) and (5)), and (c) the evaluation on government's measures against the Northern Territories dispute ("How do you evaluate the government's measures against the Northern Territories dispute?" (0 to 10 points)) (columns (3) and (6))."Δ TV view" is the difference in the viewing time of either Public Channel 1 (columns (1) through (3)) or Private Channel 4 (columns (4) through (6)) before and during the experiment (in hours). The instrument variable is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, and trust and accuracy measures.

Table 6: Difference in closenesses.

	Public_1 - Private TVs	p-value
Interacted with duration		
Japan - USA	0.225	0.025
Japan - Russia	1.133	0.000
Without interaction		
Japan - USA	0.015	0.000
Japan - Russia	0.031	0.000

Notes: For each program, the difference in closenesses to either government (domestic or foreign) is computed. Then their averages are compared between two types of broadcasters (Public Channel 1 or private channels). The difference and the associated p-value are reported. The top panel takes into account the duration of the program, while the bottom panel does not. We use machine learning to compute the closeness measures, where window size is set to 5 and vector size is set to 200. See the main text for more details.

Table 7: Effect on difference in closenesses to official statements.

	Dependent variables: Average Exposure					
	Δ Japan-USA			Δ Japan-Russia		
	(1) All	(2) P.M.	(3) Russia	(4) All	(5) P.M.	(6) Russia
<u>Public_1</u>						
Treatment	0.034	0.022	0.011	0.037	0.025	0.016
	(0.014)**	(0.007)***	(0.005)**	(0.015)**	(0.009)***	(0.007)**
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	0.35	0.13	0.08	0.40	0.18	0.13
R^2	0.74	0.66	0.60	0.77	0.65	0.63
Observations	4156	4156	4156	4156	4156	4156
<u>Private_4</u>						
Treatment	−0.005	−0.008	−0.007	0.008	−0.009	−0.007
	(0.013)	(0.008)	(0.005)	(0.016)	(0.009)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	0.19	0.16	0.10	-0.10	0.14	0.10
R^2	0.30	0.50	0.44	0.03	0.46	0.42
Observations	4156	4156	4156	4156	4156	4156

Notes: Robust standard errors are in parentheses. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$. Dependent variables are difference in closenesses to official statements made by domestic and foreign governments. We use machine learning to compute the closeness measures, where window size is set to 5 and vector size is set to 200. The top panel uses programs in Public Channel 1, while the bottom panel uses programs in Private Channel 4. First three columns use difference in the average closeness to official statements made by the Japanese and American governments during Prime Minister Abe's visit to Pearl Harbor with the former President Barack Obama. Next three columns use difference in the average closeness to official statements made by the Japanese and Russian governments during President Vladimir Putin's visit to Japan for the summit meeting. Columns (1) and (4) include all programs. Columns (2) and (5) restrict the programs to those related to PM, while columns (3) and (6) restrict them to those related to Russia. "Treatment" is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, trust and accuracy measures, and the viewing time of respective channel before the experiment.

Table 8: Effect on political evaluations of difference in closenesses to official statements.

	Dependent variables: Evaluation (North. Terr.)					
	(1) All	(2) P.M.	(3) Russia	(4) All	(5) P.M.	(6) Russia
OLS Estimates						
Δ Japan-USA	0.291	0.375	0.794			
	(0.105)***	(0.213)*	(0.332)**			
Δ Japan-Russia				0.140	0.208	0.329
				(0.103)	(0.155)	(0.216)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	4.41	4.41	4.41	4.41	4.41	4.41
R^2	0.12	0.12	0.12	0.12	0.12	0.12
Observations	4156	4156	4156	4156	4156	4156
IV Estimates						
Δ Japan-USA	7.871	12.384	24.994			
	(4.094)*	(5.715)**	(13.402)*			
Δ Japan-Russia				7.357	10.929	17.338
				(3.800)*	(5.544)**	(9.532)*
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of control	4.41	4.41	4.41	4.41	4.41	4.41
F-stat.	6.29	10.21	5.69	6.21	6.78	5.18
Effect. F-stat.	6.29	10.21	5.69	6.21	6.78	5.18
CI (AR)	[2.522, 23.919]	[4.010, 34.787]	[7.482, 77.530]	[2.391, 22.255]	[3.246, 32.661]	[4.883, 54.702]
CI (Wald)	[-.152, 15.895]	[1.182, 23.585]	[-1.274, 51.262]	[-.092, 14.806]	[.063, 21.795]	[-1.344, 36.020]
Observations	4156	4156	4156	4156	4156	4156

Notes: Robust standard errors are in parentheses. * indicates p <0.10, ** p <0.05, and *** p <0.01. Weak instrument robust inference (Anderson-Rubin (AR) test) is conducted and the corresponding CIs are reported in a row labeled as "CI (AR)". We use Stata's command `rivtest` for computing the test statistic. Effective F-stat. is the effective F-statistic of Olea and Pflueger (2013). We use Stata's `weakivtest` for computing the test statistic. Dependent variables are the evaluation on government's measures against the Northern Territories dispute. The main independent variables are difference in closenesses to official statements made by domestic and foreign governments. We use machine learning to measure the closeness. "Δ Japan-USA" is difference in the average closeness between TV programs and official statements which were made by the Japanese and American governments during Prime Minister Abe's visit to Pearl Harbor with the former President Obama. "Δ Japan-Russia" is difference in the average closeness to official statements made by the Japanese and Russian governments during President Putin's visit to Japan for a summit meeting. Columns (1) and (4) include all programs. Columns (2) and (5) restrict the sample to programs that are related to PM, while columns (3) and (6) restrict to programs that are related to Russia. The instrument is the treatment dummy. Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, trust and accuracy measures, and the viewing time of Public Channel 1 before the experiment.

Figure 1: Timeline of the experiment.

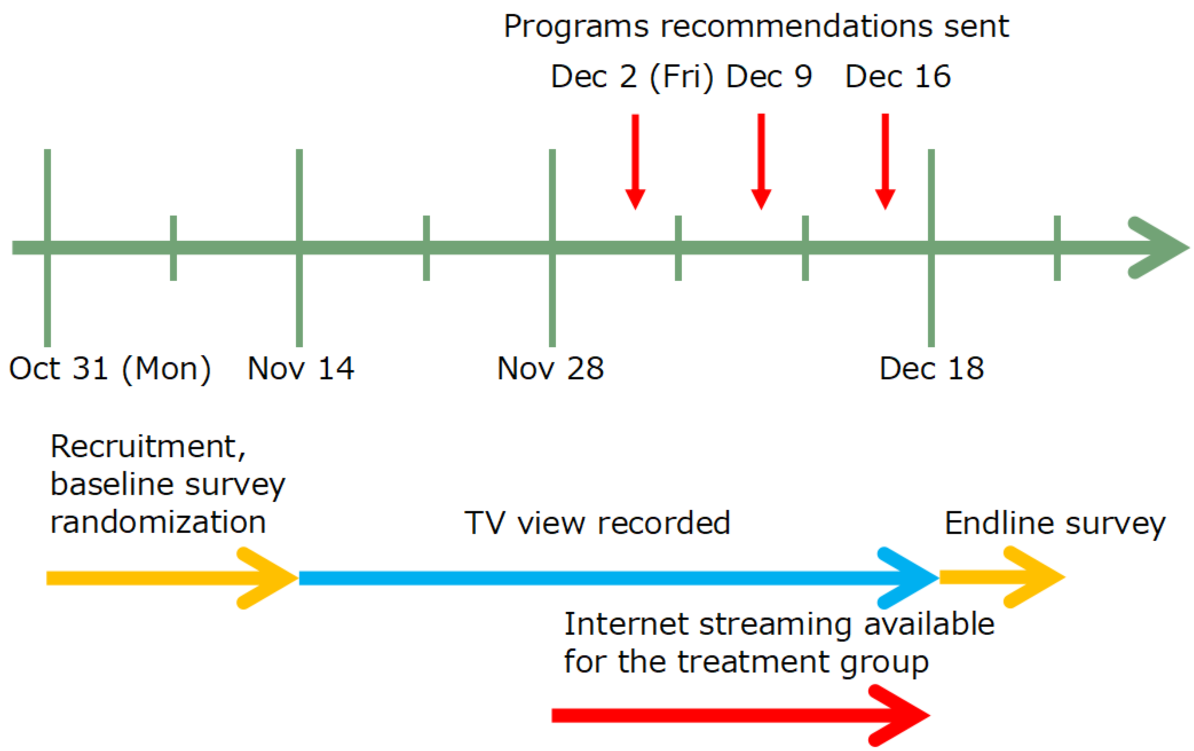
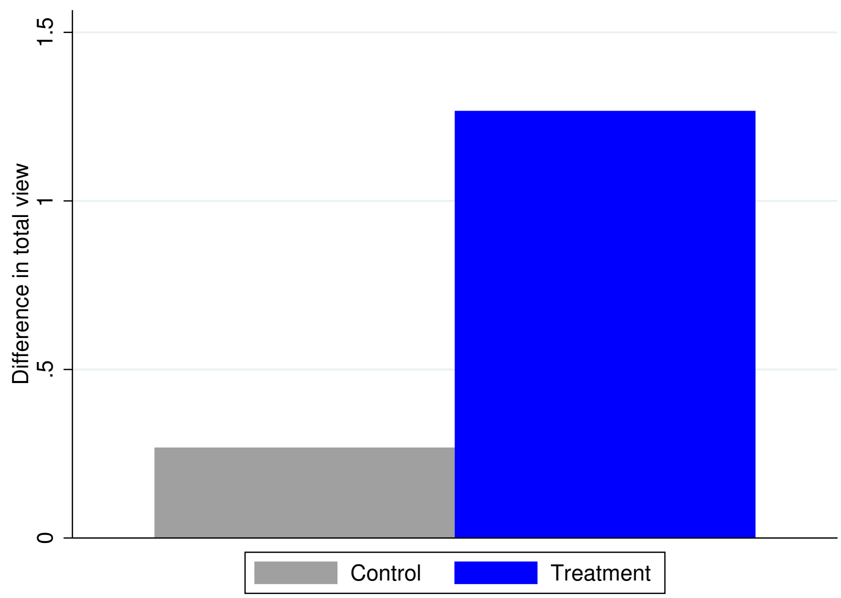
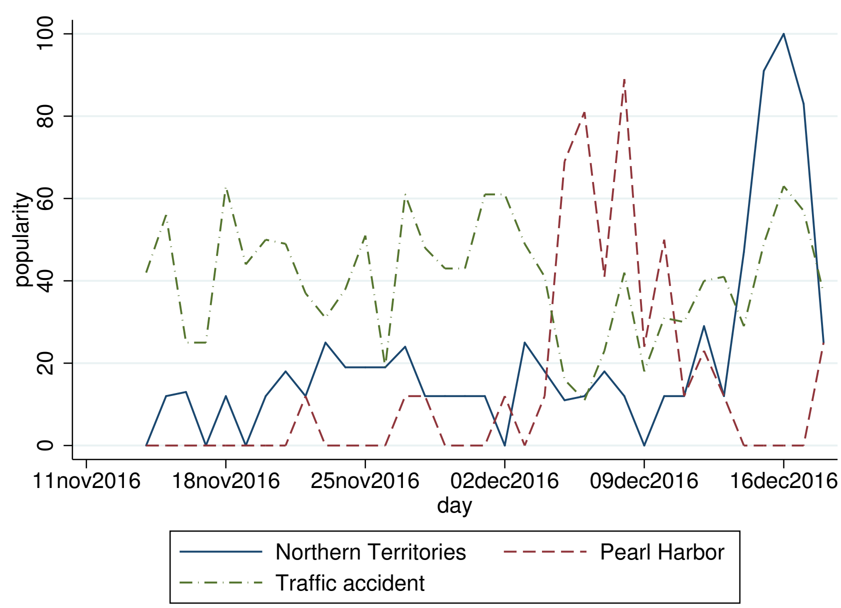


Figure 2: The average viewing time of Public Channel 1, by treatment status.



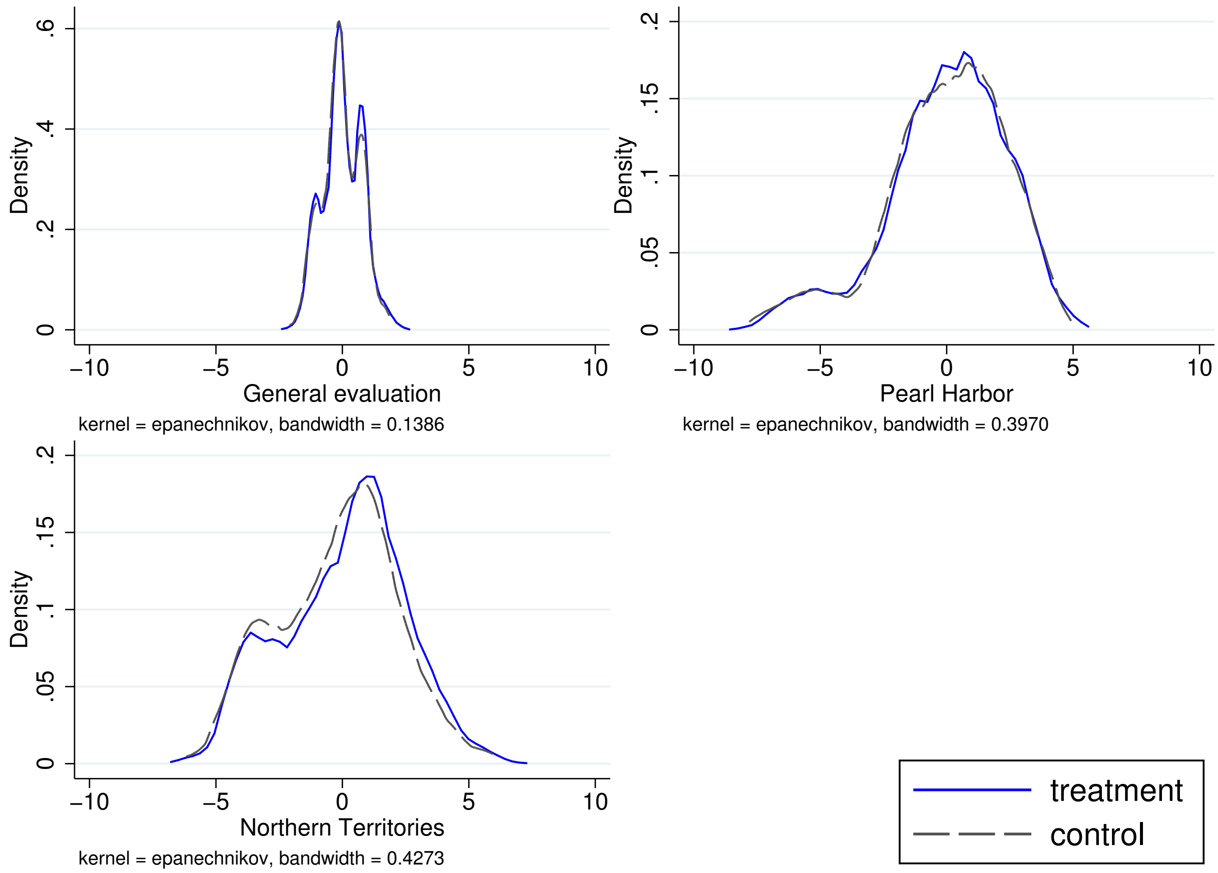
Notes: Difference in viewing time of Public Channel 1 before and during the experiment are averaged for the treatment and control groups (in hours).

Figure 3: Popularity of each topic.



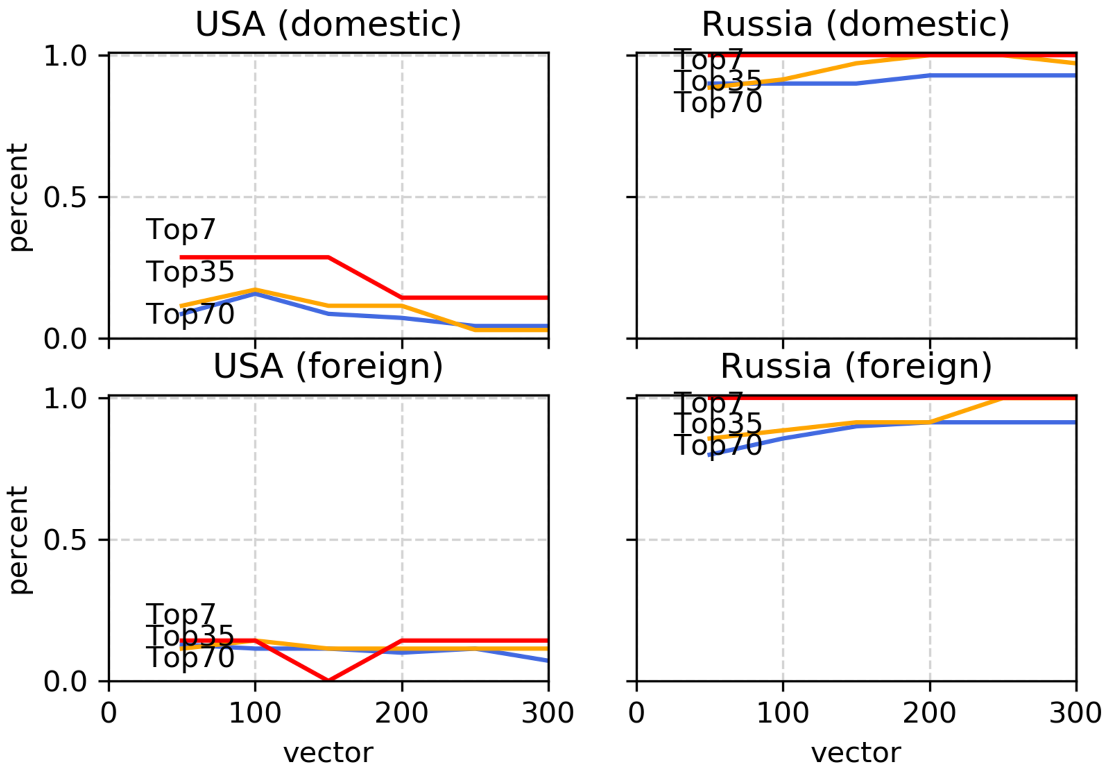
Notes: Popularity as measured by the frequency of web search using Google Trends. The location is set as Japan and the category is set as news. "北方領土" (Northern Territories) is used for Northern Territories, "真珠湾" (Pearl Harbor), and "交通事故" (traffic accident) is used for traffic accident.

Figure 4: Distributions of political outcomes: treatment versus control.



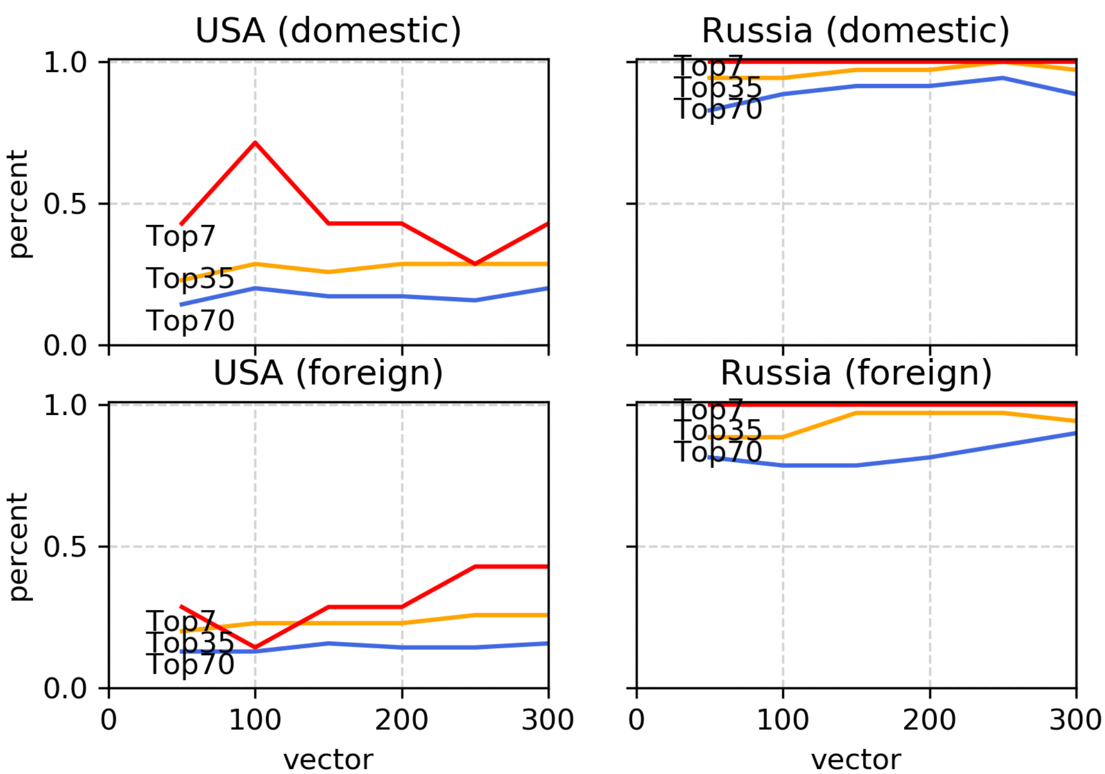
Notes: The kernel density plots of residuals from regressions of dependent variables on control variables. The dependent variable for the upper-left figure is the evaluation on government's capability ("Do you evaluate that the current government has capability to implement policies?" (4: Yes, 3: More of less, 2: Not quite, 1: No)). The dependent variable for the upper-right figure is the evaluation on Prime Minister Abe's visiting Pearl Harbor with President Obama ("How do you evaluate Prime Minister Abe's visiting Pearl Harbor with President Obama to console the souls of war victims?" (0 to 10 points)). The dependent variable for the bottom-left figure is the evaluation on government's measures against the Northern Territories dispute ("How do you evaluate the government's measures against the Northern Territories dispute?" (0 to 10 points)). Control variables are a dummy for supporting the Liberal-Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading Sankei newspaper, and trust and accuracy measures.

Figure 5: Predictability of our machine, USA and Russia topics (window=15).



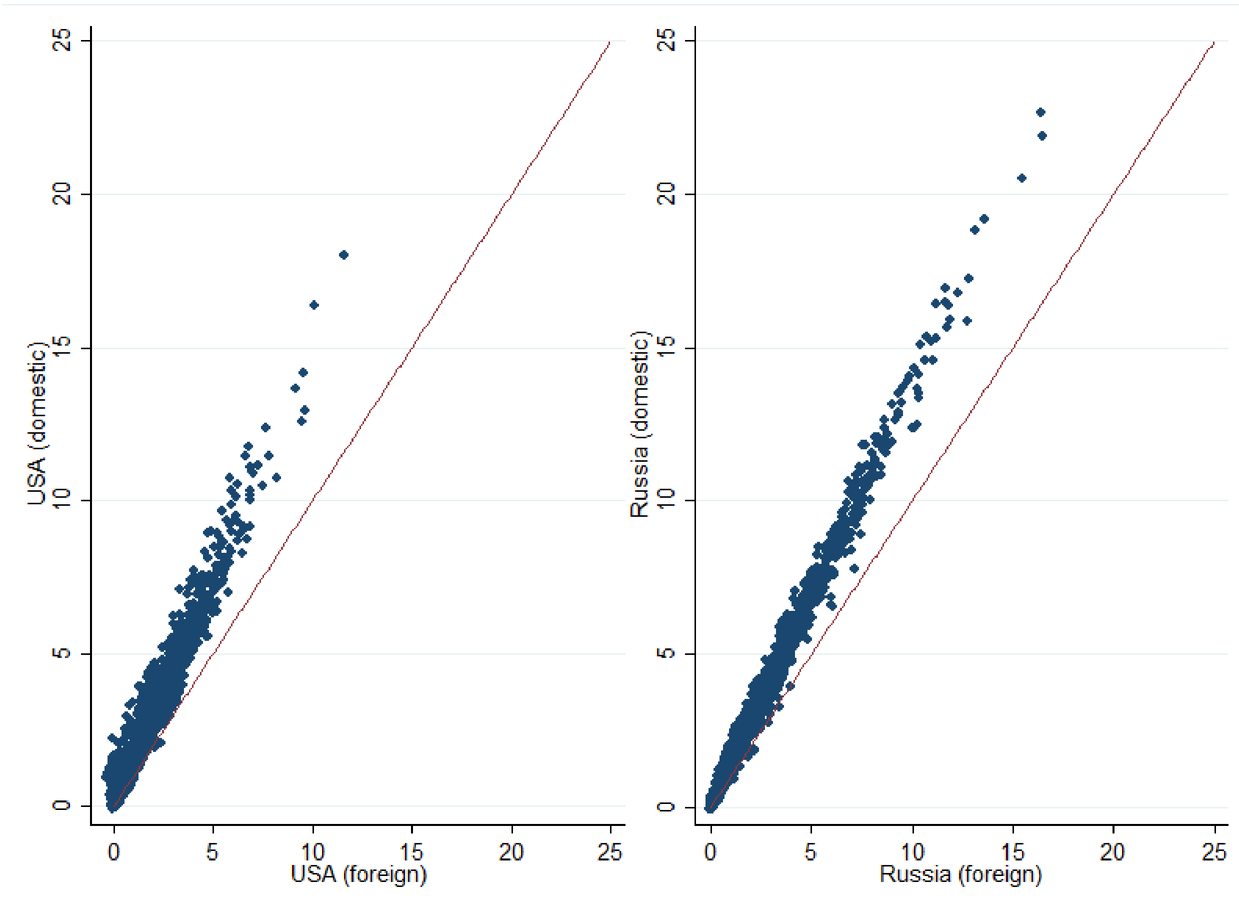
Notes: The figures plot the percentage of TV programs that are categorized as either the USA topic or the Russia topic among the programs that are the closest to official statements (top7, top35, and top70): the left figures use closeness to the Japanese government's statement (top) and the American government's statement (bottom), respectively during Prime Minister Abe's visit to Pearl Harbor with President Obama; the right figures use the Japanese government's statement (top) and the Russian government's statement (bottom), respectively during President Putin's visit to Japan for the summit meeting. We use machine learning to make the closeness measures, in which window size is set to 15. See the main text for more details.

Figure 6: Predictability of our machine, USA and Russia topics (window=5).



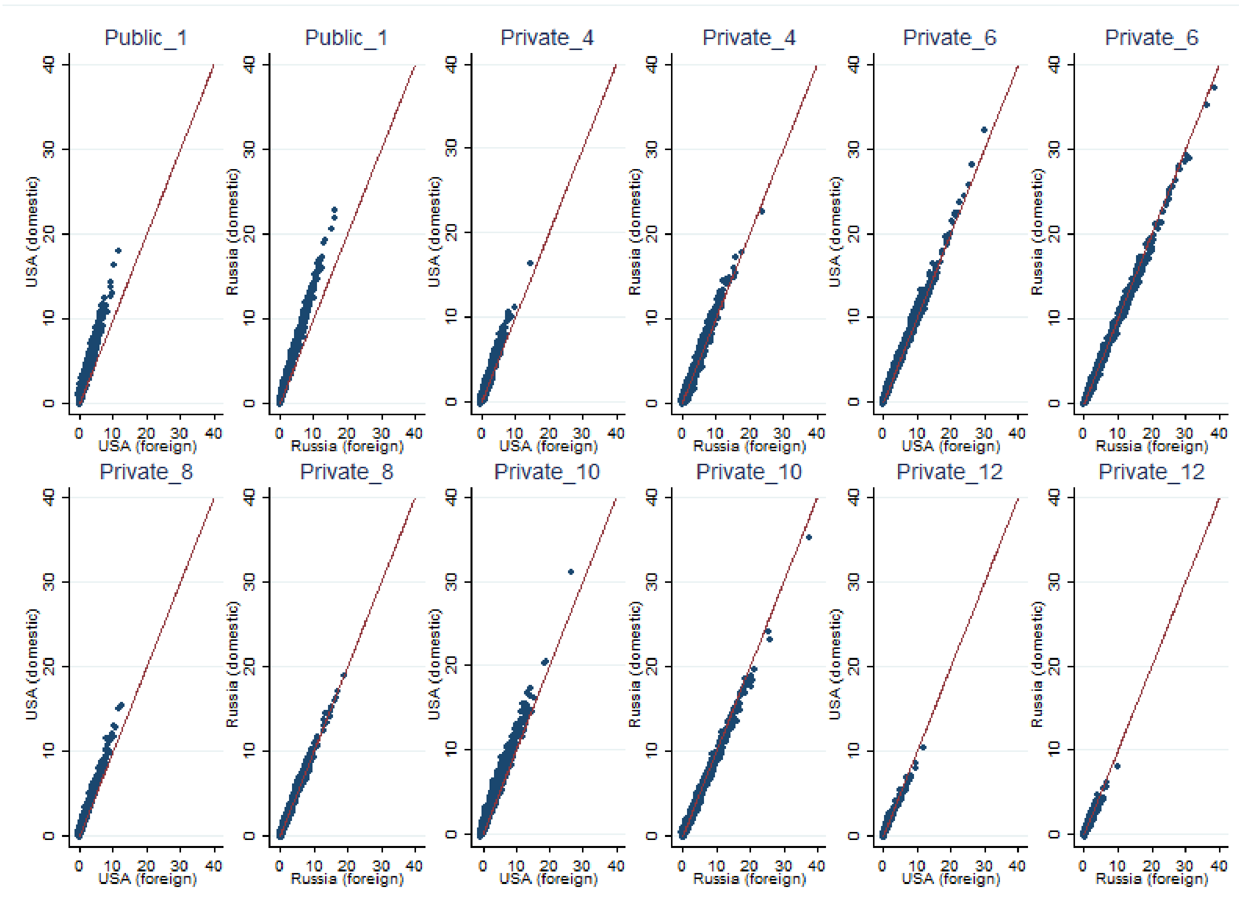
Notes: The figures plot the percentage of TV programs that are categorized as either the USA topic or the Russia topic among the programs that are the closest to official statements (top7, top35, and top70): the left figures use closeness to the Japanese government's statement (top) and the American government's statement (bottom), respectively during Prime Minister Abe's visit to Pearl Harbor with President Obama; the right figures use the Japanese government's statement (top) and the Russian government's statement (bottom), respectively during President Putin's visit to Japan for the summit meeting. We use machine learning to make the closeness measures, in which window size is set to 5. See the main text for more details.

Figure 7: Average closeness to official statements interacted with individuals' viewing time, domestic versus foreign (Public Channel 1).



Notes: The figures plot the closeness to official statements, interacted with individuals' viewing time (AvgExposure in (3)). The figure shows the results for Public Channel 1. In the left panel, closeness to the Japanese government's statement with interaction (y-axis) is plotted against the American government's statement interacted with interaction (x-axis). In the right panel, the Japanese government's statement with interaction (y-axis) is plotted against the Russian government's statement interacted with interaction (x-axis). We use machine learning to make the closeness measures, in which window size is set to 5 and vector size is set to 200. See the main text for more details.

Figure 8: Average closeness to official statements interacted with individuals' viewing time, domestic versus foreign (private broadcasters).



Notes: The figures plot the closeness to official statements, interacted with individuals' viewing time (AvgExposure in (3)). The figure shows the results for private broadcasters, except for the top-left figures, which shows the results for Public Channel 1 for comparison. For each channel, in the left panel, closeness to the Japanese government's statement with interaction (y-axis) is plotted against the American government's statement interacted with interaction (x-axis), while in the right panel, the Japanese government's statement with interaction (y-axis) is plotted against the Russian government's statement interacted with interaction (x-axis). We use machine learning to make the closeness measures, in which window size is set to 5 and vector size is set to 200. See the main text for more details.