PUBLIC MEDIA DO SERVE THE STATE: A FIELD EXPERIMENT*

Shuhei Kitamura† Toshifumi Kuroda‡

November 26, 2019

Abstract: A vital role of public media according to the theoretical literature is to enable consumers to cross-check the information provided by slanted private media. To examine how the information provided by public media differs from that of private media, this study analyzes the slant and persuasive effect of public media in a democracy. In 2016, we conducted a randomized field experiment in Japan in collaboration with the nation's public service broadcaster in which the capacity for viewing its programs was randomly increased. We find that the treatment increased the evaluation of certain policies by raising the viewing time of its programs. To study the mechanism behind these findings, we develop a novel method to measure media slant using the semantic similarity between TV programs and the official statements made by domestic and foreign governments. To implement this method, we collected TV scripts from all TV channels in the Tokyo region using a TV digital recorder. Using this metric, we find that public media's information is semantically closer to the domestic government's statements than to those of the foreign government compared with private media's information. Moreover, this positive effect is driven by subjects being exposed to such slanted information during the experiment. By contrast, the effect is not explained by the competition between political parties. These findings raise the importance of creating institutions that secure the independence of public media.

JEL codes: K23, L82, D78

*This study was financially supported by JSPS grant numbers 16K03663 and 18K12768 as well as the Japan Broadcasting Corporation. The views and opinions in this paper do not necessarily represent those of the 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 their helpful comments and suggestions. We are also grateful to the seminar participants at GRIPS, Keio, Osaka, Osaka City, and Rochester and at the EEA-ESEM Annual Meetings at the University of Manchester, 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. (Word count: 12,493 words)

†Osaka School of International Public Policy, Osaka University, 1-31 Machikaneyama, Toyonaka, Osaka, 560-0043 Japan. Phone: +81 6-6850-5645. Fax: +81 6-6850-5656. Email: kitamura@osipp.osaka-u.ac.jp.

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

1 Introduction

Many countries have public media outlets including the BBC in the United Kingdom, PBS in the United States, and SVT and SR in Sweden. Their mission includes providing public services through media outlets such as television and radio to consumers. Moreover, public media are the most trusted outlets in many countries including Germany, Japan, and the United Kingdom, according to a study by the Reuters Institute at Oxford University (Newman et al., 2018). PBS is also the most trusted media outlet by both Republicans and Democrats in the United States (Pennycook and Rand, 2019). One of the reasons for this high degree of trust might be that public media are believed to have no particular political position compared with private media.

A growing literature on media slant confirms its existence 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 outlets 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, empirical studies of public media are limited. Is public media's information different from that of private media? If so, how and why? Is it similarly persuasive? Given that public media outlets tend to co-exist with private media outlets in many countries, studying public media's information and influence should thus complement previous studies of private media.

More importantly, theories indicate that the equilibrium information of media should change depending on the existence of public media in the market. Assuming that public media's information is less slanted than private media's, the existence of such media outlets should reduce the degree of media slant in the market by enabling consumers to cross-check information (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006).[1] For example, although Republicans (Democrats) may not watch MSNBC's (Fox News') programs, they

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

may still be willing to watch PBS's programs. If PBS's programs are less slanted, consumers, in theory, should not easily be persuaded by the information delivered by private media because they can cross-check their own camp's information with that of PBS.

By contrast, if public media's information is also somewhat slanted, this might affect public opinion more easily because, as just described, public media tend to be the most trusted outlets in many countries. It thus becomes crucial to check whether the neutrality assumption holds for public media and whether they can indeed persuade consumers.

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

Using our experiment, which is designed to randomly increase the capacity to view the PSB's TV programs, we estimate the causal effect of watching the PSB's programs on policy evaluations and the government's capability to implement policies. We find that the treatment increased individuals' evaluations of certain policies by raising the viewing time of programs. By contrast, we find no effect on the general evaluation of the government's capability to implement policies.

To examine the mechanism behind these findings, we use a novel method to measure media slant in which we apply unsupervised machine learning to measure the semantic similarity between the content of TV programs and that of official statements made by domestic and foreign governments. We employ the method proposed by Le and Mikolov [2014], which is a general text classification method using the fixed-length feature representations of pieces of texts. Specifically, we first train the machine so that it learns the semantic similarity between words using TV scripts as a corpus (i.e., a list of plain texts such as Google Books Corpora). Then, we measure the semantic similarity between TV programs and official

[2] These features are similar to those of the BBC in the United Kingdom.

statements based on the words included in them.

Our method has two advantages. First, it does not require researchers to select words of interest *ex ante*. Previous studies have often used predetermined word categories created by scholars (Tetlock, 2007; Qin et al., 2017, 2018). Using these methods, however, it is often difficult for researchers to select the relevant words (e.g., finding party-specific words for political candidates in the multi-party system). Moreover, the process itself can be time-consuming. By contrast, our method does not involve selecting words before computing the semantic similarity between them. Rather, the machine computes the semantic similarity using a plain corpus.

Hence, as long as a relevant corpus is used, we can, in theory, measure the semantic similarity between *any* documents such as the official documents of authorities and international organizations for which it is difficult to find specific words to represent them. For example, it might be difficult to find words that represent an international organization such as the United Nations (UN). However, our method allows us to measure how much a specific (possibly non-UN) document is semantically closer to the UN document without specifying words that would represent the UN *ex ante*.

Second, compared with similar methods employed by 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 differences across documents.[3] Regarding the last point, roughly speaking, our algorithm employs a "fixed-effect" procedure in which the estimated parameter captures the contextual specificity of a document when maximizing the probability of observing a target word.

To the best of our knowledge, this is the first study to apply this method in the context of social sciences. Since it can, in theory, be used to measure the semantic similarity between any documents, it has many applications in economics and other subjects in social sciences.

Using this similarity measure, we find that the PSB's information is closer to the domestic

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

government's statements than to those of the foreign government compared with private media's information. Moreover, individuals' being exposed to such information during the experiment resulted in their positive evaluations of certain policies.

This study contributes to a large body of the literature on media and information in economics and political science. First, it contributes to the literature on the source of media slant, which can be driven by either supply-side factors (e.g., owners, advertisers, and journalists) or demand-side factors (e.g., consumers). Besley and Prat [2006] show that an incumbent is more likely to "buy" unfavorable information from the media outlet in order not to disclose information to voters if this is less costly than the rents from holding office. Further, empirical studies find that media controlled by a state, party, or 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).

By contrast, in a more democratic setting, namely in the United States, Gentzkow and Shapiro [2010] show that media slant in newspapers is driven by consumer preferences rather than owner preferences. Gentzkow et al. [2015] also find little evidence that the ruling party influences the partisan composition of the press. These mixed findings in different settings suggest that factors affecting media slant, either supply-side or demand-side factors, may depend on the degree of media independence and market competition in the media market.

On the contrary, even in a democratic setting, our study indicates that public media can be more susceptible to media capture than private media. In other words, supply-side factors seem to drive public media's slant in democracies. We argue later that this might be because consumers must pay mandatory receiving fees.

In addition, our study complements previous studies of persuasion by private media that find the positive impact of a 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). By contrast, another study shows that slanted newspapers have been found to have no influence on the vote share, political knowledge, or political opinions, although reading such newspapers does increase support for Democratic candidates and raise voter

turnout (Gerber et al., 2009). Similar to the first strand of the literature, we find that public media also have a positive impact on political outcomes.

The literature on the impact of new information technology on political outcomes is also related to this work because our experiment involves online streaming. Previous studies show that voter turnout has fallen because of the diffusion of TV (Gentzkow, 2006) and the Internet (Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019). Using US data, Prior [2005] finds that new information technology has widened the knowledge gap and turnout between individuals who prefer news and those who prefer entertainment. By contrast, Boxell et al. [2017] find that the greater use of the Internet is not associated with political polarization in the United States.

Finally, our study contributes 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; Ash and Labzina, 2019). As mentioned above, our method has several advantages including no requirement for selecting words of interest *ex ante*.

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

2 Background

2.1 Market Structure

The television market in Japan has a dual structure that consists of one nationwide PSB and several regional private broadcasters. The PSB has two channels, 1 and 2. In addition, local private channels in each region 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 Japan. In the greater Tokyo area, for example, there are five major private channels: 4, 6, 8, 10, and 12. The last entry or exit of private broadcasters in this area occurred in 1995.

The PSB's market share varies from region to region as the number of private broadcasters varies (from two to six). In the greater Tokyo area in 2016, the market share of Channel 1 was 18.4% compared with Channels 4 (22.5%), 6 (16.0%), 8 (15.5%), 10 (19.8%), and 12 (7.8%) according to Dentsu Media Innovation Lab [2018].

2.2 Source of Revenue and Ownership

The PSB is mostly (99% of annual income) funded by the receiving fees paid by households and enterprises that have TV equipment, regardless of whether they actually view the PSB's programs or not. It also receives a relatively small (0.5% of annual income in 2016) government subsidy for providing international and party election broadcasts. Any revenue from advertising is prohibited by the Broadcast Act, which also rules that the PSB's annual budget must be approved by the Diet.

The 12 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.[4] The appointment sometimes comes to public attention. For example, the appointment in 2013 gained media attention because some members were too closely related to Prime Minister Abe (*Nikkei*, 2013). Historically, all members have been Japanese nationals.

By contrast, the main source of revenue for private broadcasters is advertising (82.4% of their annual income in 2016) according to Japan Commercial Broadcasters Association [2017]. Foreign nationals can buy the private broadcaster's share, although the Broadcast Act regulates that foreign ownership with voting rights should not exceed 20%.[5]

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

[5] If foreign ownership without voting rights is counted, the share of foreign shareholding can exceed 20%.

2.3 Program Contents

Regarding the editing of broadcasting programs, the Broadcast Act rules that the broadcaster "shall not harm public safety or good morals," "shall be politically fair," "shall not distort the facts," and "shall clarify the points at issue from as many angles as possible where there are conflicting opinions concerning an issue" (Ministry of Internal Affairs and Communications, 2010). Although the same rule applies to all broadcasters, there are some differences in program content between public and private outlets. For example, Krauss [2000] finds that the frequency at which journalists and newscasters express their personal opinions about a news item in private broadcasting programs is twice that in the PSB's programs. In addition, when PSB newscasters comment on a particular news, they tend to provide 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.[6] This style of the PSB's news programs might be one of the major reasons for this.

3 Experiment

3.1 Overview

An online randomized controlled trial was conducted in the greater Tokyo area in November and December 2016.[7] As the main treatment, we randomly provided free subscriptions to view the PSB's programs online.[8] In particular, treated individuals could view the PSB's programs on web browsers and mobile phones. The services comprised live streaming and

However, in practice, it does not deviate too much from this threshold.

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

[7] 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 it to provide an Internet streaming service if it did not harm market competition. To test whether the streaming service met this criterion, the PSB conducted large-scale randomized experiments in 2015 and 2016; this study focuses on the 2016 experiment. Regarding the 2015 experiment, Kuroda et al. [2017] find no evidence that the streaming service harmed market competition. The 2016 experiment was significantly larger than the 2015 experiment.

[8] In the second experiment, we also sent program recommendations to randomly selected subgroups in the treatment group on Fridays. The Online Appendix provides the analyses of this second treatment.

video on demand (VOD). The treatment thus increased the 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.[9] Some programs were not available during the experiment because of copyright issues.[10] In addition, because of government regulation, only programs that started at 7:00 am or after and ended by 11:00 pm were available for both services. By contrast, the VOD service was accessible at any time.

We recruited 6,000 Internet users using a mail survey.[11] First, we created 40 blocks according to age (20s, 30s, 40s, 50s, and 60s), sex (male and female), and TV viewing frequency (almost every day, at least once a week, at least once a month, and hardly ever or never). Then, we recruited individuals so that the percentage of each block in the sample was proportional to the population in each category in the greater Tokyo area.[12]

Recruitment was conducted by 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 explain the outline of the surveys and rewards.[13] The email clarified that the purpose was to survey the individual's daily TV viewing behavior including private programs, daily media access, and media evaluations, particularly TV. The email also stated that participants would keep a viewing record every day. To minimize selection bias, the availability of free Internet streaming services was not informed at the time of recruitment. Participants were told about this benefit as a surprise just before the start of the experiment. Finally, 5,000 of the 6,000 individuals were randomly assigned to the treatment group and the remainder were assigned to the control group.[14]

[9] The monthly subscription fee for the VOD service was 972 JPY (about 10 USD) or a minimum of 100 JPY (about 1 USD) per program.

[10] About 16% of total broadcasting time was not provided in these services.

[11] The number was decided to maximize the statistical power given the budget restrictions and government regulation.

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

[13] The company did not disclose the details of the rewards to us, but they were based on its incentive criterion, according to the company's reply. Most importantly, there was a bonus for completing a viewing record (see the Data section).

[14] We used the `randomize` command in Stata 14. We allocated more than half of the individuals to the

Figure I shows the timeline of the experiment. We started recruitment at the end of October 2016. The baseline survey was conducted at the same time. We then randomized the subjects. All participants started to keep a viewing record from November 14. The free online services became available for the treatment group on November 28. Finally, the endline survey was conducted after the end of the experiment on December 18.

3.2 Data

Our main data are each subject's viewing record. Participants received an email from the survey company every day in which they were asked to mark the TV programs (the PSB's and private broadcasters' programs) that they viewed on that day. Although the data were self-reported, we alleviated measurement errors in a number of ways. First, to reduce the reporting cost, participants simply check-marked the names of the TV programs they viewed on the computer screen. Second, participants were incentivized for providing reports and were given extra time before submitting their reports.[15] Third, unreliable samples were dropped as a robustness check. Finally, the instrumental variable (IV) method was employed. As explained above, recording started two weeks before the start of the experiment.[16]

In the baseline survey, we asked individuals' socioeconomic status, media usage, media evaluations, and political orientation. In the endline survey, we asked for their evaluations of policies and the government's capability in addition to media usage and media evaluations.

In addition, we collated unique data: TV scripts from all TV channels in the Tokyo region including private broadcasters. We collected these data using a TV digital recorder.[17]

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 decided not to participate

treatment group because of the second experiment, as described and analyzed in the Online Appendix.

[15] This bonus was given to individuals at the end of the experiment. The survey company did not tell participants the exact amount or formula to avoid strategic behavior. Participants were given three days to complete their reports.

[16] For treated individuals, the access log of the PSB's Internet streaming service was also available. We could not use the access log for the main analysis because there was no such data for the control group. See the Online Appendix for our analysis using these data for the subset of individuals.

[17] The machine provided 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.

in the experiment, and an individual's viewing records were deemed to be unreliable,[18] our final sample became 4,453 (74.2%).[19]

We conducted balance checks between the treatment and control groups in the final sample using variables from the baseline survey (see the results in the Online Appendix). Most variables are balanced between the two groups, suggesting that the randomization was successful. Since some variables were not balanced,[20] we included them as the baseline control in the following analysis. Table I shows the summary statistics of the 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, Channel 1 (henceforth Public 1). In particular, we run the following regression using ordinary least squares (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 1 before and during the experiment, Treatment_i is the treatment dummy, $\mathbf{X_i}$ is a vector of the control variables, and ε_i is the error term, which we assume is independently distributed across individuals.[21] Our main interest is the estimate of β. Using the difference in viewing time, we remove any systematic differences in the usual viewing pattern across individuals that we cannot perfectly control.

Figure II shows the average of the dependent variable for the treatment and control

[18] The individual reported that he/she viewed most TV programs in a day.
[19] The Online Appendix also provides an analysis for checking attrition bias.
[20] These variables were a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, and trust and accuracy measures.
[21] Recall that there is a pre-treatment period in the viewing record from which we obtain information on pre-treatment viewing behavior.

groups. It shows that the treatment group viewed Public 1 approximately one hour longer than the control group during the experiment.

Table II shows the average treatment effect on the viewing time of Public 1. The first column does not include the baseline control variables, while the next two columns do include them. The third column drops individuals who occasionally failed to report their viewing records for certain programs, although they were likely to view them, according to the access log. Overall, the treatment increased the viewing time of Public 1 by about one hour, corresponding to Figure II. In the following analysis, we use the smallest sample to minimize measurement errors.

4.2 Effect by Week and on Other TV Channels

Table III shows the effect by week.[22] The effect appears in all three weeks of the experiment. Moreover, the size of the effects varies little across these weeks, although the effect in Week 2 is slightly larger than those in the other weeks. This finding shows that the treatment effect does not simply capture the effect of special daily programs.

Next, to understand the overall impact of our treatment, we investigate the effect on other TV channels. We are particularly interested in knowing whether the increase in the viewing time of Public 1 is explained by a decrease in the viewing time of another channel (Channels 2, 4, 6, 8, 10, and 12), holding total viewing time constant, or an increase in total viewing time.

Columns (1)–(7) in the top panel of Table IV show the average treatment effects by channel. The dependent variables use the difference in viewing time as in equation (1), but for each channel. Column (1) replicates column (3) in Table II for comparison purposes. As shown in column (2), the treatment did not increase the viewing time of Channel 2 (henceforth Public 2), another channel broadcast by the PSB. Public 2 is slightly different from the other channels, as it focuses on educational programs. We also find that it was the

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

least popular channel in our data.[23] These factors might explain why the treatment did not increase the viewing time of Public 2.

Columns (3)–(7) show the treatment effects for all the private broadcasters. We find that the treatment effects for Channels 4 and 12 are negative, although the latter effect is relatively weak. By contrast, the last column of the top panel in Table IV 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 1 is to some extent explained by the change in the viewing time of private channels, holding total viewing time constant. Using the difference in shares instead of the difference in hours for the dependent variables, the bottom panel of Table IV shows a consistent picture; the treatment also increased the share of Public 1 and decreased the share of Channel 4. The effect on Channel 12 now comes out as insignificant.

4.3 Effect on Political Attitudes

Next, we investigate the effect of watching Public 1 on political attitudes. Two major political events related to foreign policy occurred during the experiment. The first was President Vladimir Putin's visit to Japan on December 15 and 16 to attend a summit. One of the main topics of the meeting was the Northern Territories dispute, a territorial dispute between Japan and Russia over four 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). The second major event was that Prime Minister Shinzō Abe announced that he would visit Pearl Harbor with then-President Barack Obama. Although the event actually occurred after the experiment (December 26–28), the first public announcement was made on December 5 (i.e., during the experiment).

Figure III shows the popularity of each topic as measured by the frequency of Google

[23] The average viewing time of Public 2 during the experiment was only about one hour, while that of Public 1 was about 11 hours, as shown in Table I.

searches in Japan according to Google Trends. It shows that "真珠湾" (Pearl Harbor) became more popular following the first public announcement by Prime Minister Abe and that "北方領土" (Northern Territories) became more popular during the summit. For comparison purposes, the search results for "交通事故" (traffic accident) were also presented, which showed no systematic pattern.

We included survey questions on these two major political events in the endline survey: How do you evaluate the government's measures against the Northern Territories dispute and how do you evaluate Prime Minister Abe's visit to Pearl Harbor with President Obama. Individuals rated each question on a scale of 0 (the lowest rating) to 10 (the highest rating). In addition, we asked individuals to evaluate the capability of the government to implement policies, with four possible answers from agree to disagree.

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

$$\text{Evaluation}_i = \phi + \psi \Delta\text{Public_1}_i + \mathbf{X_i}\xi + v_i, \tag{2}$$

for individual i, where Evaluation_i is the political evaluation just described, $\Delta\text{Public_1}_i$ is the difference in the viewing time of Public 1 as in equation (1), $\mathbf{X_i}$ is a vector of the baseline control variables, and v_i is the error term, which we assume is independently distributed across individuals. We are interested in the estimate of ψ.

Simply regressing model (2), the estimated ψ may be biased because $\Delta\text{Public_1}_i$ is likely to be endogenous. This 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 thus use the treatment variable as the instrument for $\Delta\text{Public_1}_i$ in the IV framework.

Columns (1)–(3) in Table V show the results. The top panel shows the OLS estimates of equation (2). The results are not statistically significant for all the dependent variables.

Next, the middle panel reports the IV (two-stage least squares, 2SLS) estimates in which the treatment variable is used as the instrument. Although watching Public 1 longer affects neither the evaluation of the government's capability nor the evaluation of Prime Minister Abe's visit to Pearl Harbor (columns (1) and (2), respectively), it has a statistically significant positive effect on the evaluation of 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 the resulting 95% confidence intervals and p-values. Column (3) shows that the confidence intervals include only positive values and that the effect remains significant.

The estimate (0.243) implies that a one standard deviation increase in the viewing time of Public 1 (8.848) increases the evaluation by 2.15, or approximately a half of the control mean. Hence, this effect is relatively large. Overall, the OLS estimates are underestimated, suggesting the second channel of bias as described above.[24]

Finally, the bottom panel shows the estimates of the reduced-form regressions. As expected, the intention-to-treat effect on the evaluation of the government's capability and Prime Minister Abe's visit to Pearl Harbor is null, while that on the evaluation of the Northern Territories dispute is statistically different from zero and positive.

Figure IV plots the probability distribution of these three outcomes for the treatment and control groups. The dependent variables are regressed on the baseline controls and the residuals are then plotted to draw these figures. As expected, the distributions for the two outcomes (the evaluation of the government's capability and Prime Minister Abe's visit to Pearl Harbor) are the same. By contrast, the distributions for the Northern Territories dispute show that the evaluations have shifted to the right for the treatment group compared with the control group. Moreover, the difference is observed over a wide range of domains,

[24] Another interpretation might be that the average treatment effect is smaller than the local average treatment effect. However, given that the independent variable is most likely to be endogenous, this is less likely to be the main reason.

suggesting that the effect is not particularly driven by individuals who have certain prior opinions on that issue.

4.3.1 Effect of Not Viewing Private TV Programs

As shown earlier, Table IV indicates some substitution between Public 1 and Channel 4. In this section, we analyze whether the effect on political evaluations also comes from *not* viewing Channel 4. Columns (4)–(6) in Table V use the difference in the viewing time of Channel 4 instead of Public 1 as the independent variable. The top panel shows the OLS estimates, while the middle panel shows the IV (2SLS) estimates for which the treatment dummy is used as the instrument.

According to the table, the OLS estimates for the evaluation of the government's capability to implement policies and on Prime Minister Abe's visit to Pearl Harbor are positive and statistically significant; however, these estimates are most likely to be biased because of endogeneity. By contrast, the IV estimates are significant only for the evaluation of the government's measures against the Northern Territories dispute, although the effect is relatively weak. Moreover, the sign of all the IV estimates becomes negative. Because the F-statistic is smaller than the conventional level, 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 IV that the treatment seems to have *decreased* the viewing time of Channel 4, the negative sign indicates that the positive evaluation of the Northern Territories issue may also come from *not* viewing Channel 4 to some degree. However, in the next section, we present suggestive evidence that the result of the positive evaluation is most likely due to viewing Public 1 more rather than viewing Channel 4 less.

5 Mechanism Behind the Findings

In this section, we use TV script data to investigate the mechanism behind these findings. In particular, we analyze information on the TV programs to which individuals were exposed

during the experiment to understand why the treatment had a significant positive effect on the evaluation of certain policies.

5.1 Closeness to Official Statements

Given that our main outcomes concern foreign policies, we hypothesize that treated individuals might have been exposed more to information more closely related to that provided by the domestic government than by the foreign government.

To test this hypothesis, we apply an unsupervised machine learning algorithm to compute the semantic similarity between TV programs and the official statements made by domestic (Japan) and foreign governments (United States and Russia). Then, we calculate the sum of the similarity measure between each TV program and official document, weighted by the duration of each program, for each individual. In other words, this variable captures the average exposure to information provided by either government. Finally, exposure to the foreign government's information is subtracted from exposure to the domestic government's information to formulate a *relative* exposure variable.

Formally, for individual i and official statement $j \in \{\text{Japan, USA, Russia}\}$, 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 that 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 a value between -1 (very far) and 1 (very close). The higher the value of Closeness_{pj}, the more similar are program p and government j's official statement. We use machine learning to construct this closeness variable.[25]

Next, we take the difference between these measures for the domestic government and

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

foreign government $k \in \{\text{USA, Russia}\}$:

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

Intuitively, $\Delta\text{AvgExposure}_{ik}$ aims to capture the average exposure to the official statement made by the domestic government relative to the average exposure to the official statement made by the foreign government. This variable takes a higher positive value if individuals are exposed more to information more closely related to that provided by the domestic government than the foreign government.

5.2 Construction of the Closeness Variable

To construct the variable Closeness_{pj} in (3), we apply an unsupervised machine learning algorithm that learns fixed-length feature representations from pieces of texts (Le and Mikolov, 2014).

To obtain the semantics of words from unlabeled data, Firth [1957] initially proposed the distributional hypothesis, which assumes that words used in the same contexts tend to have similar meanings. In our case, the similarity of words is captured by the distributed semantics, which assumes that the meaning of a word is represented by fixed-length numerical vectors.[26]

We train a model such that it maximizes the probability of observing a word given a context (e.g., predict the fourth word from "the," "cat," and "sat").[27] Once the process is finished, we compute the similarity between each official statement and each TV program broadcast during the experiment using the semantic similarity between the words in these documents. We use the average of the word semantics in a document to compute the semantics of that document. Using this definition, two documents are semantically more similar if they contain more words that are semantically more similar.

[26] The distributional hypothesis and distributed semantics are summarized in, for example, Eisenstein [2019].

[27] This is also called the paragraph vector with distributed memory model.

Let $\mathscr{D} = \{D_1, ..., D_K\}$ be a set of documents (TV programs) used to train a machine and $\mathbf{X}$ be a vector of all the words in a training corpus, with length n. Each document D_k contains a sequence of words. Let $x_i \in \mathbf{X}$ be a target word and $\mathbf{c}_i^k \in \{0,1\}^{n\times 1}$ be the local context $x_{i-s}, ...,$ x_{i-1}, $x_{i+1}, ...,$ x_{i+s} of the target, where s is the maximum window size, given the sequence of words 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 and zero otherwise.

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

Using a softmax function, the probability of observing the target word x_i in document D_k given the local context $\mathbf{c}_i^k$ and memory $\mathbf{d}^k$ is

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

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

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

Given the trained $\mathbf{P}$, one can compute $\mathbf{d}^k$ by taking the average of the columns in $\mathbf{P}$ of the words included in document D_k (Rong, 2016).

Suppose two documents each contain a single sentence: "I have a cat" and "I have a dog". Thus, $\dim \mathscr{D} = 2$ and $\mathbf{X} = (\text{dog}, \text{cat}, \text{a}, \text{have}, \text{I})^T$ without loss of generality. Let the window size and vector size be $s = 1$ and $h = 2$, respectively.

The sentence "I have a cat" has a sequence of words: "I," "have," "a," and "cat." The local context of "a," given the window size 1, is therefore "have" and "cat." Hence, c_3^1 is written as $(0,1,0,1,0)^T$. Similarly, the local context of "a" in the sentence "I have a dog" has $c_3^2 = (1,0,0,1,0)^T$. These local contexts of the "a"'s are projected onto a two-dimensional

hidden space using matrix P and then onto an output using matrix Q. These matrices are constantly updated in the training process so that "a" can be predicted with a high probability.

When the training is finished, words with similar semantics can be mapped to a similar location in the vector space. For tractability, one can typically evaluate the similarity between words x_i and x_j using cosine similarity $\frac{\mathbf{q}_{x_i} \cdot \mathbf{q}_{x_j}}{\|\mathbf{q}_{x_i}\| \times \|\mathbf{q}_{x_j}\|}$.

Applying the same method to measure the semantic similarity between documents, we compute the semantic similarity between documents D_k and D_l using $\frac{\mu_k \cdot \mu_l}{\|\mu_k\| \times \|\mu_l\|}$, where μ_m is the average of each element of the q_x's of the words contained in document D_m, treating duplicates as separate words. Therefore, we assume that the semantics of a document are represented by the average semantics of words in that document. We check this assumption when we set the parameter values.

As described in the Introduction, a major advantage of the algorithm is that it takes into account the ordering and semantics of words as well as the 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 the *ex-ante* selection of words by researchers. Instead, it computes the semantic similarity between words using only a corpus.

The data used to train the machine (i.e., a corpus) are the scripts of TV programs broadcast after the experiment until August 28, 2018.[28] This choice of the corpus is based on the characteristics of our study; other studies may also use other corpora such as Google Books Corpora depending on their purpose. We use Python's `gensim` package for all the procedures.

Since Japanese texts, in contrast to western texts, are not separated into single words in the original data, we parse the scripts using a common Japanese parser called `MeCab`. We

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

use only nouns, verbs, adjectives, and adverbs, so that our measures capture the similarity between texts in a meaningful way.[29] Moreover, we apply stemming for all words. The number of TV programs in the training data is 195,016, which contain 613,497 distinct words.

For the official statements, we use those made by the Japanese and US governments when Prime Minister Abe and President Obama visited Pearl Harbor[30] and those by the Japanese and Russian governments when Prime Minister Abe and President Putin met at the summit in Japan.[31] 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, the only difference between the official statements for each event is whether the statement was made by either the domestic government or the foreign government.

Before starting, the algorithm requires researchers to decide two 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 obtain the baseline numbers for these parameters, we compute the percentage of the closest programs to the official statements that can be categorized as either the United States or Russia topic to see how much we can successfully predict the relevant programs.

In particular, we first list the TV programs and order them from the most to the least close to each statement for each event. We then select several programs from the closest programs and check the percentage that can be categorized as either the United States or

[29] The following parts of speech are excluded: interjection, conjunction, particles, auxiliary verbs, prefix, and determiner.

[30] The official statements in Japanese are available on 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).

[31] The official statements for both countries are only available in Japanese from the Cabinet Office's website (https://www.kantei.go.jp/jp/97_abe/statement/2016/1216kaiken.html).

Russia topic.[32] Thus, a higher percentage means that we are more successful at predicting the relevant programs for each topic. We choose the top 7, 35, and 70 closest programs, which roughly correspond to 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 V plots a figure for a window size of 15 and Figure VI plots a figure for a window size of 5. In both, the left figures use the closeness to the Japanese government's statement (top) and the US government's statement (bottom). By contrast, the right figures use the closeness to the Japanese government's statement (top) and the Russian government's statement (bottom).

These figures show that the machine with a window size of 15 is worse at predicting programs for the United States topic, particularly when the closeness to the US government official statement is used. In addition, Figure VI, which uses a window size of 5, shows that the percentage roughly decreases as the vector size increases for the top 7 programs if the closeness to the Japanese official statement is used, while it increases for the top 7 programs if the closeness to the US official statement is used, with the lowest hit when the vector size is 100.

Based on these analyses, we select a window size of 5 and a vector size of 200 for the main analysis. The Online Appendix provides the results using the other parameter values.

5.3 Results

Table VI shows the difference in the closeness between the programs and official statements without taking into account participants' viewing behavior (i.e., $\text{Duration}_p \times \text{Closeness}_{pj}$ in (3)). The table is used to check for any difference in information from the supply side. We first take, for each program, the difference in the closeness to either government (domestic or foreign) and then compare the averages between the two types of broadcasters (Public 1 and private channels). The top panel of the table also takes into account the duration of the program, while the bottom panel does not.

First, the table shows that all the differences are statistically significant. Moreover, they

[32] See the Online Appendix for the procedure of categorizing programs into these topics.

are all positive, indicating that Public 1's information is closer to the domestic government's than to that of the foreign government compared with private broadcasters' information. Interestingly, the difference between Japan and Russia is somewhat larger than that between Japan and the United States.

Next, we take into account participants' actual viewing behavior. Each point in Figure VII corresponds to the individual's exposure to particular information, namely AvgExposure_{ij} in (3). Exposure to the domestic government's information is plotted against exposure to the foreign government's information. The red lines indicate the 45-degree lines.

Strikingly, the left panels show that the data points tend to be located above the 45-degree lines, especially for individuals who are away from the origin, indicating that they were more likely to be exposed to information closer to the domestic government's information than that of the foreign government by viewing Public 1's programs. This result becomes even starker when compared with the right panels, which plot the same figures but for the private channels. This shows that the data points are more or less on the 45-degree lines even though there are five times more data points in the panels compared with the respective left panels. These results suggest that individuals were affected by information somewhat unique to Public 1 during the experiment.

To check whether the positive second-stage result on the political evaluations is explained by this informational difference, we next use the closeness measures in the regressions. The top panel of Table VII shows the first-stage results in which the treatment dummy is used as the main independent variable and the similarity between Public 1 and official statements is used as the dependent variable. "Δ Japan–USA" means the difference in average closeness between the Japanese and US governments, while "Δ Japan–Russia" is the difference in average closeness between the Japanese and Russian governments.

Columns (1) and (4) include all TV programs broadcast by Public 1, columns (2) and (5) use only programs categorized as the topic related to Prime Minister Abe, and columns (3) and (6) use only programs categorized as the Russia topic. Overall, the table shows that the treatment group was more likely to receive information closer to that provided by the

domestic government than by the foreign government. This finding may not be surprising given the slant in Public 1's information found before.

Next, the top panel of Table VIII shows the OLS estimates in which the evaluation of the Northern Territories dispute is used as the dependent variables (i.e., the same dependent variables used in column (3) in Table V) and the closeness measures are used as the independent variables. The bottom panel of the 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 VII; the dependent variable uses either all programs or a subset of them focusing on a particular topic depending on the column.

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

In terms of the magnitude, using the estimate in column (1) (7.389), we find that a one standard deviation (0.681) increase in the independent variable increases the evaluation of the Northern Territories dispute by 5. Similarly, the estimate in column (4) (7.357) shows that a one standard deviation increase in the treatment variable increases the evaluation by 5. Therefore, the effect size is considerable.

Because of the (albeit weak) second-stage effect through Channel 4 shown in the previous section, we also examine whether the effect on the political evaluation also comes from *not* being exposed to information closer to that provided by the foreign government. To check this possibility, we formulate closeness measures for Channel 4 using the same procedure. The bottom panel of Table VII shows the first-stage estimates of regressing these measures on the treatment dummy. The table does not present clear evidence that the treatment

induced individuals *not* exposed to information closer to that of the foreign government. Compared with the top panel in the same table, the estimates are also relatively small. This is not surprising given that individuals do not seem to be exposed to the slanted information through Channel 4, as shown in Figure VII.

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

Overall, we find that treated individuals positively evaluated certain foreign policies because they were exposed more to information closer to that provided by the domestic government during the experiment.

6 Discussion

Why is public media's information closer to that provided by the domestic government than the foreign government compared with private media's information? As noted in the Introduction, the literature shows that media slant is driven by either demand-side factors (Gentzkow and Shapiro, 2010) or supply-side factors (Enikolopov et al., 2011; Durante and Knight, 2012; Qin et al., 2018; Chen and Yang, 2019). Regarding the demand-side factors, since receiving fees are collected regardless of viewers' political orientation, it is hard to believe that consumers' political preferences (e.g., right vs. left) drive the slant in public media's information. The same argument holds for public media in other countries such as Sweden and the United Kingdom in which similar fees or taxes are collected from most adults.

Turning to the supply-side factors, since receiving fees are mandatory for anyone who has TV equipment, changing information in public media should not 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 advertising.

These factors indicate that the 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 toward the ruling party than the opposition party compared with private media's information. We test this hypothesis in the Online Appendix, but find no results supporting this hypothesis, although we do find that public media's information is slanted toward the ruling party. In other words, our results do not seem to be explained by the 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, only the result for the Northern Territories dispute in Table V is significant. A possible explanation might be that the summit had just finished when the endline survey was conducted, while Prime Minister Abe's visit to Pearl Harbor had not. The former event might thus have endorsed individuals' beliefs and hence influenced their survey answers. An alternative explanation could be that our closeness measures are not perfect for eliciting the information to which individuals were exposed during the experiment despite being the most intuitive measures available.

7 Conclusion

In contrast to the literature on the slant and persuasion of private media, this study examines those of public media in a democracy. To estimate the causal effect of public media, we conducted a randomized field experiment in collaboration with the nation's PSB in Japan in which the capacity for viewing its TV programs randomly increased. We find that the treatment increased the evaluation of the government's foreign policies by raising the viewing time of programs.

To understand the mechanism behind these findings, we develop a novel method in which we apply machine learning to measure media slant using the semantic similarity between the content of TV programs and the official statements made by the domestic and foreign

governments. The method has several advantages including no requirement for selecting words *ex ante* to measure the semantic similarity. This means that the method can be applied to cases that lack clear representative words. Using the similarity measure, we find that treated individuals were exposed more to information closer to that provided by the domestic government than that provided by the foreign government during the experiment, which explains the positive evaluation of the policies.

According to the theoretical literature, a vital role of public media is to enable consumers to cross-check the information provided by slanted private media (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006). The findings of the present study that public media's information is slanted, and, even worse, has a persuasive effect thus raise the importance of creating institutions that secure the independence of public media.

Although we conducted a randomized experiment to obtain a generalizable causal effect, the study was conducted in a specific country. Thus, further analysis of the role of public media in other countries may be needed. In addition, future studies should address the mechanism of how humans change their beliefs and attitudes by, for example, closely studying the neural response to information.

References

Ash, E. and E. Labzina, "Does Politicized Media Distort Political Discourse? Evidence from U.S. Cable News Channels," Mimeo, 2019.

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

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

Campante, F. R., R. Durante, and F. Sobbrio, "Politics 2.0: The Multifaceted Effect of

Broadband Internet on Political Participation," Journal of the European Economic Association, 16 (2018), 1094–1136.

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

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

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

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

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

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

Eisenstein, J., Introduction to Natural Language Processing, 1st ed. (Cambridge, MA: MIT Press, 2019).

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

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

Firth, J. R., Papers in linguistics, 1934-1951, 1st ed. (London: Oxford University Press, 1957).

Gavazza, A., M. Nardotto, and T. M. Valletti, "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies," Review of Economic Studies, 86 (2019), 2092—2135.

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

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

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

— — "What Drives Media Slant? Evidence from U.S. Daily Newspapers," Econometrica, 78 (2010), 35–71.

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

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

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

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

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

Krauss, E. S., Broadcasting Politics in Japan: NHK and Television News, 1st ed. (Ithaca: Cornell University Press, 2000).

Kuroda, T., D. Kido, and Y. Ichikawa, "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, 2017.

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

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

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

Ministry of Internal Affairs and Communications, The Broadcast Act, 2010 (http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Resources/laws/pdf/090204_5.pdf).

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

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

Nikkei, NHK Keiei Iin Kokkai Doui, 2013 (https://www.nikkei.com/article/DGXNASDD080D5_Y3A101C1TJ0000/).

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

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

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

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

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

— —, "Media bias in China," American Economic Review, 108 (2018), 2442–2476.

Rong, X., "word2vec Parameter Learning Explained," Mimeo, 2016.

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

Strömberg, D. "Media Coverage and Political Accountability: Theory and Evidence," in Handbook of Media Economics, Volume 1, S. P. Anderson, J. Waldfogel, and D. Strömberg, ed. (Amsterdam: North Holland Publishing Co., 2015).

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

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

Tables and Figures

Table I: Summary statistics

	Mean	Std. Dev.	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
Channel_4	25.23	33.38	0.00	413.88	4452
Channel_6	17.03	23.02	0.00	309.90	4452
Channel_8	16.77	25.94	0.00	230.25	4452
Channel_10	22.26	28.68	0.00	362.97	4452
Channel_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 Channel_4	8.16	15.13	−62.15	134.75	4452
Δ Channel_6	5.18	11.27	−66.05	123.45	4452
Δ Channel_8	5.79	11.76	−56.00	113.37	4452
Δ Channel_10	7.45	13.16	−70.98	135.70	4452
Δ Channel_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 II: The average treatment effect on watching Public 1

	Dependent variable: 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 *** $p<0.01$. The dependent variables are the difference in the viewing time of Public 1 before and during the experiment (in hours). "Treatment" is the treatment dummy. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, and trust and accuracy measures. Column (3) drops individuals who occasionally failed to report the viewing records of certain programs, although they were likely to have seen them according to the access log.

Table III: The average treatment effect on watching Public 1, by week

	Dependent variable: 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 *** $p<0.01$. The dependent variables are the weekly viewing time of Public 1 during the experiment (in hours). Week 1: November 28–December 4, Week 2: December 5–11, and Week 3: December 12–18. "Treatment" is the treatment dummy. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, trust and accuracy measures, and the viewing time of Public 1 before the experiment.

Table IV: The average treatment effect on watching TV programs, by channel

	Dependent variable: Δ TV view							
	(1) Public_1	(2) Public_2	(3) Channel_4	(4) Channel_6	(5) Channel_8	(6) Channel_10	(7) Channel_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 *** $p < 0.01$. The 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. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, and trust and accuracy measures.

Table V: Effect of watching TV programs on political evaluations

	Dependent variable: Evaluation					
	Public_1			Channel_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. The dependent variables are (a) the evaluation of the government's capability ("Do you evaluate that the current government has the capability to implement policies?" (4: Yes, 3: More or less, 2: Not quite, 1: No)) (columns (1) and (4)), (b) the evaluation of Prime Minister Abe's visit to Pearl Harbor with President Obama ("How do you evaluate Prime Minister Abe's visit to Pearl Harbor with President Obama?" (0 to 10 points)) (columns (2) and (5)), and (c) the evaluation of the 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 1 (columns (1)–(3)) or Channel 4 (columns (4)–(6)) before and during the experiment (in hours). The IV is the treatment dummy. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, and trust and accuracy measures. Effective F-stat. is the effective F-statistic of Olea and Pflueger [2013]. We use Stata's `weakivtest` to compute the test statistic. Weak instrument robust inference (Anderson–Rubin (AR) test) is conducted and the corresponding p-values and confidence intervals are reported in rows labeled as "p-value (AR)" and "CI (AR)," respectively. We use Stata's command `rivtest` to compute the test statistic.

Table VI: Difference in closeness

	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 the closeness to either government (domestic or foreign) is computed. Then, their averages are compared for the two types of broadcasters (Public 1 and (the average of) private channels). The differences and associated p-values 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 the window size is set to 5 and the vector size is set to 200. See the main text for more details.

Table VII: Effect on the difference in the closeness to the official statements

	Dependent variable: Average Exposure					
	Δ Japan–USA			Δ Japan–Russia		
	(1) All	(2) PM	(3) Russia	(4) All	(5) PM	(6) Russia
Public_1						
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
Channel_4						
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 *** $p<0.01$. The dependent variables are the difference in the closeness to the official statements made by the domestic and foreign governments. We use machine learning to compute the closeness measures, where the window size is set to 5 and the vector size is set to 200. The top panel uses the programs on Public 1, while the bottom panel uses the programs on Channel 4. The first three columns use the difference in average closeness to the official statements made by the Japanese and US governments during Prime Minister Abe's visit to Pearl Harbor with President Barack Obama. The next three columns use the difference in average closeness to the official statements made by the Japanese and Russian governments during President Putin's visit to Japan for the summit. Columns (1) and (4) include all programs. Columns (2) and (5) restrict the programs to those related to Prime Minister Abe, while columns (3) and (6) restrict them to those related to Russia. "Treatment" is the treatment dummy. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, trust and accuracy measures, and the viewing time of the respective channels before the experiment.

Table VIII: Effect on political evaluations of the difference in the closeness to the official statements

	Dependent variable: Evaluation (North. Terr.)					
	(1) All	(2) PM	(3) Russia	(4) All	(5) PM	(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
p-value (AR)	0.00	0.00	0.00	0.00	0.00	0.00
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 *** $p<0.01$. The dependent variables are the evaluation of the government's measures against the Northern Territories dispute. The main independent variables are the difference in the closeness to the official statements made by the domestic and foreign governments. We use machine learning to compute the closeness measures, where the window size is set to 5 and the vector size is set to 200. "Δ Japan–USA" is the difference in average closeness between TV programs and the official statements made by the Japanese and US governments during Prime Minister Abe's visit to Pearl Harbor with President Obama. "Δ Japan–Russia" is the difference in average closeness to the official statements made by the Japanese and Russian governments during President Putin's visit to Japan for the summit. Columns (1) and (4) include all programs. Columns (2) and (5) restrict the sample to programs related to Prime Minister Abe, while columns (3) and (6) restrict to programs related to Russia. The instrument is the treatment dummy. The control variables include a dummy for supporting the Liberal Democratic Party, a minor party, and no party, a dummy for not reporting annual income and for reading the Sankei newspaper, trust and accuracy measures, and the viewing time of Public 1 before the experiment. Effective F-stat. is the effective F-statistic of Olea and Pflueger [2013]. We use Stata's `weakivtest` to compute the test statistic. Weak instrument robust inference (AR test) is conducted and the corresponding p-values and confidence intervals are reported in rows labeled as "p-value (AR)" and "CI (AR)," respectively. We use Stata's command `rivtest` to compute the test statistic.

Figure I: Timeline of the experiment

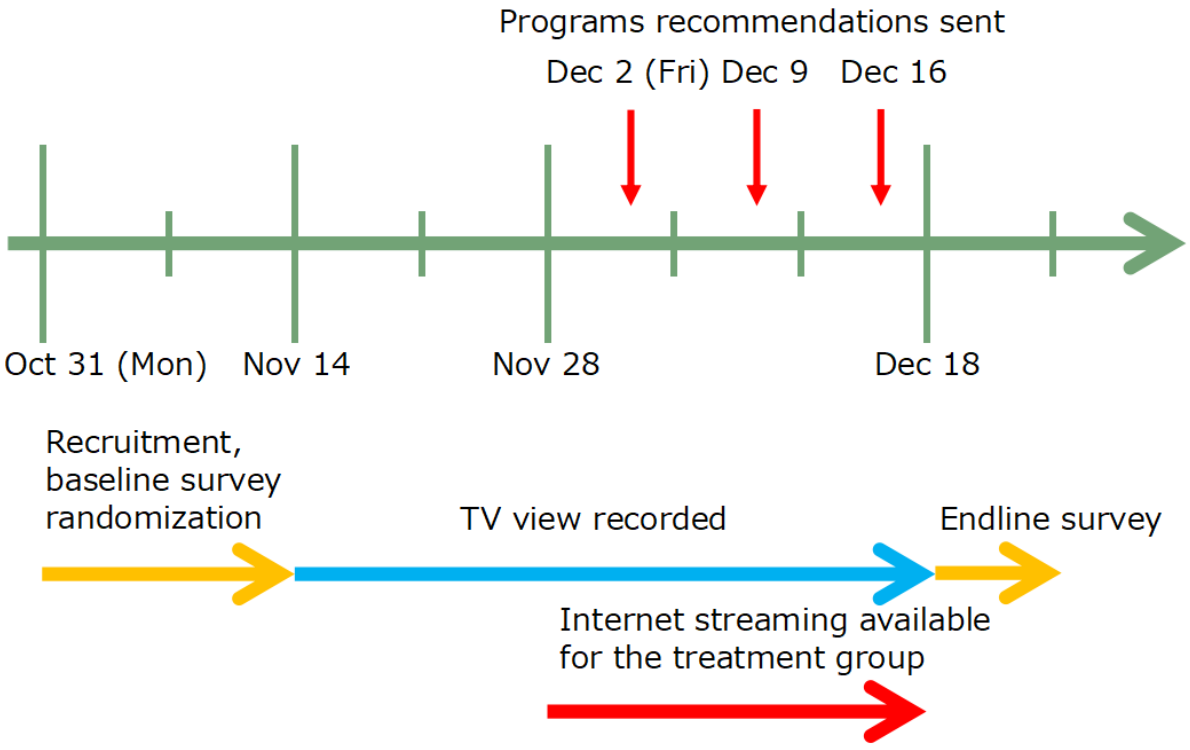
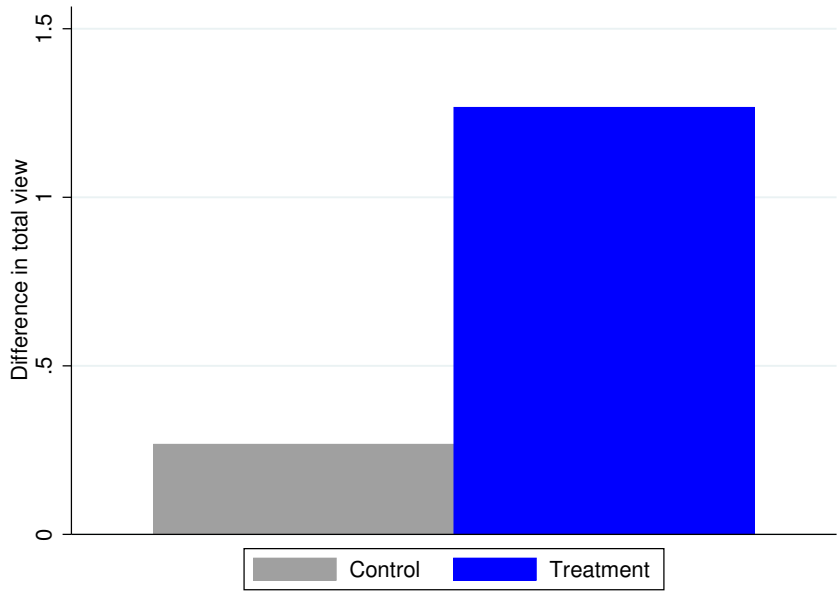
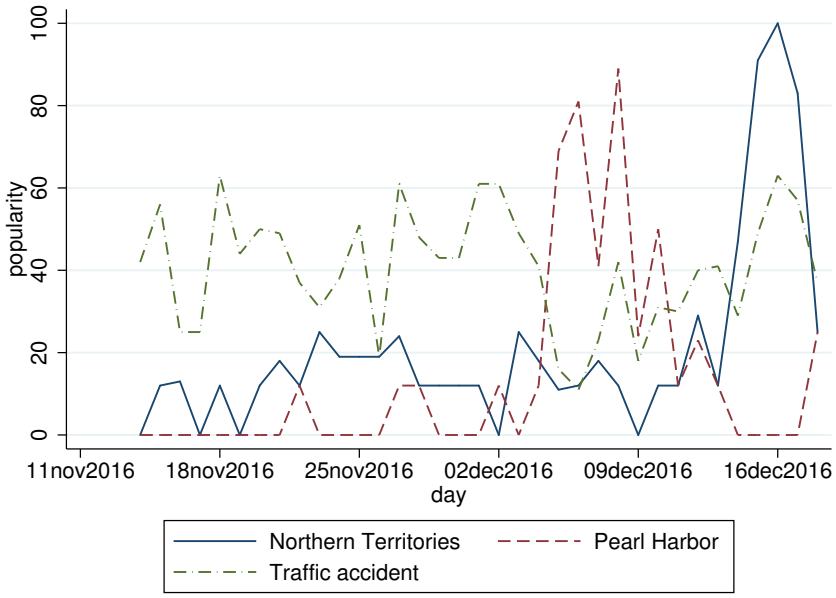


Figure II: The average viewing time of Public 1, by treatment status



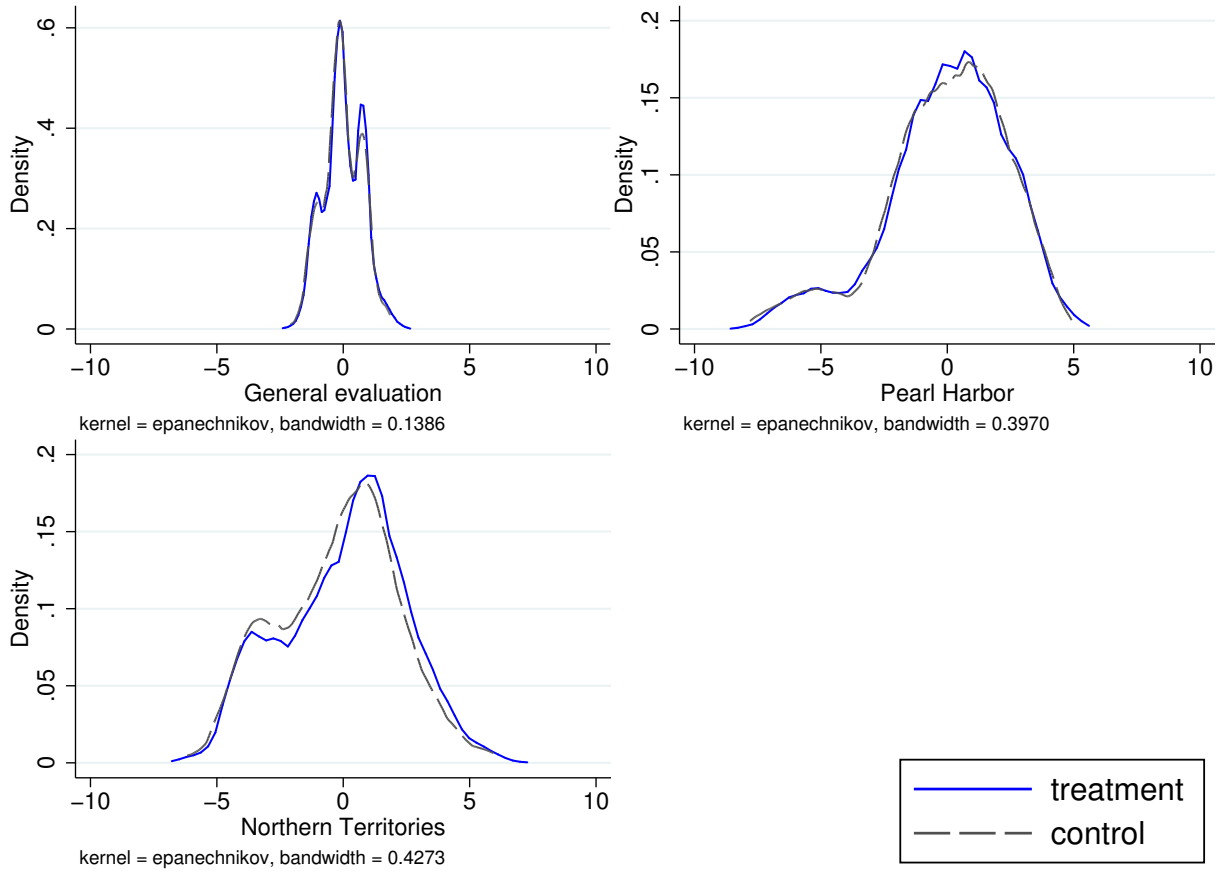
Notes: Difference in the viewing time of Public 1 before and during the experiment are averaged for the treatment and control groups (in hours).

Figure III: Popularity of each topic



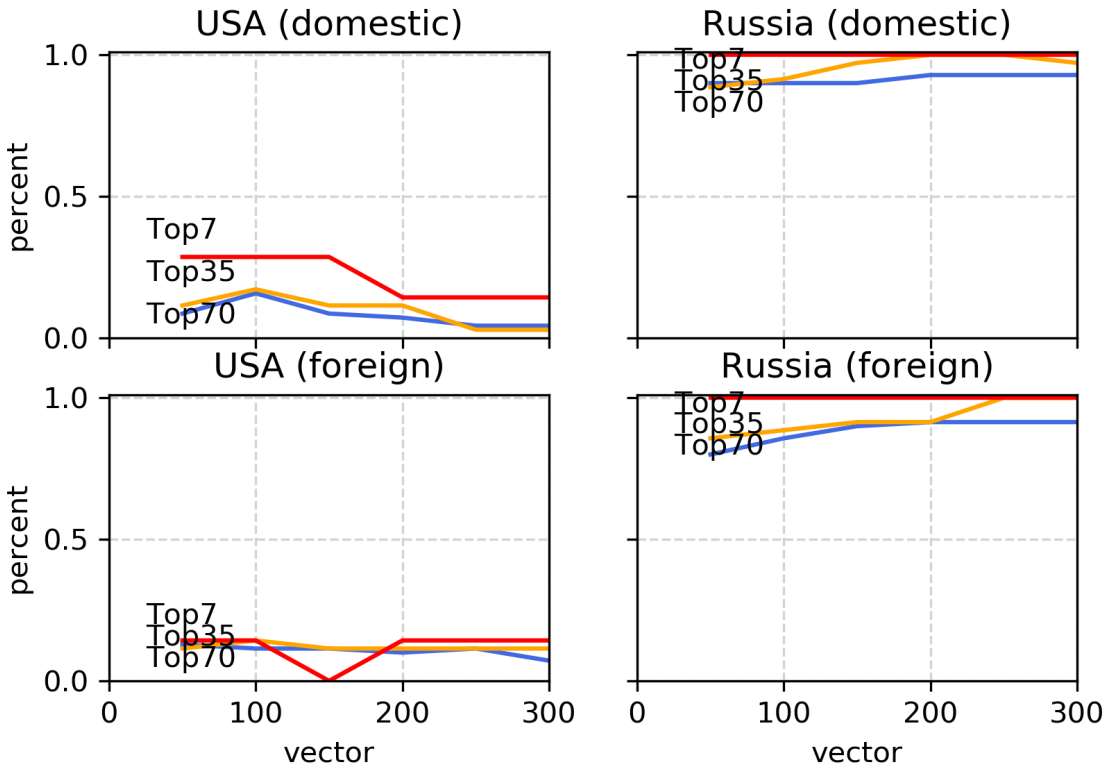
Notes: Popularity is measured by the frequency of web searches using Google Trends. The location is set as Japan and the category is set as news. "北方領土" (Northern Territories) is used for Northern Territories, "真珠湾" (Pearl Harbor) is used for Pearl Harbor, and "交通事故" (traffic accident) is used for traffic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

Figure IV: Distributions of political outcomes: Treatment versus control



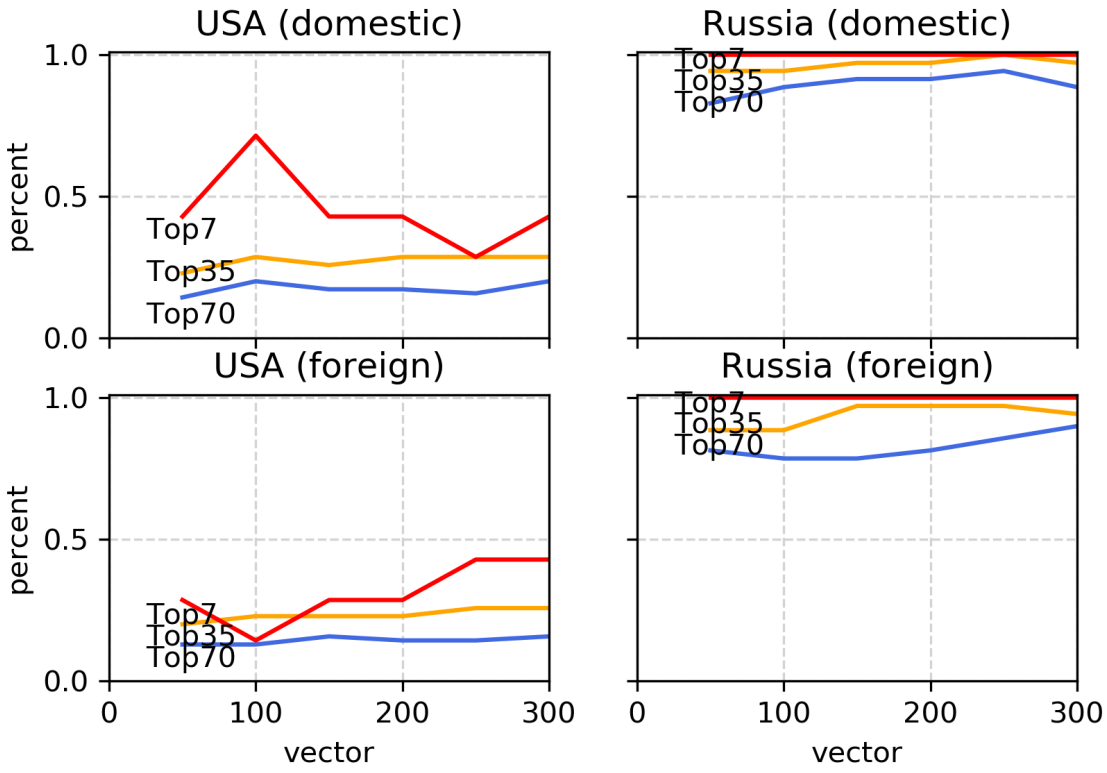
Notes: The kernel density plots of the residuals from the regressions of the dependent variables on the control variables. The dependent variable for the top-left figure is the evaluation of the government's capability ("Do you evaluate that the current government has the capability to implement policies?"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

Figure V: Predictability of the machine learning, United States and Russia topics (window=15)



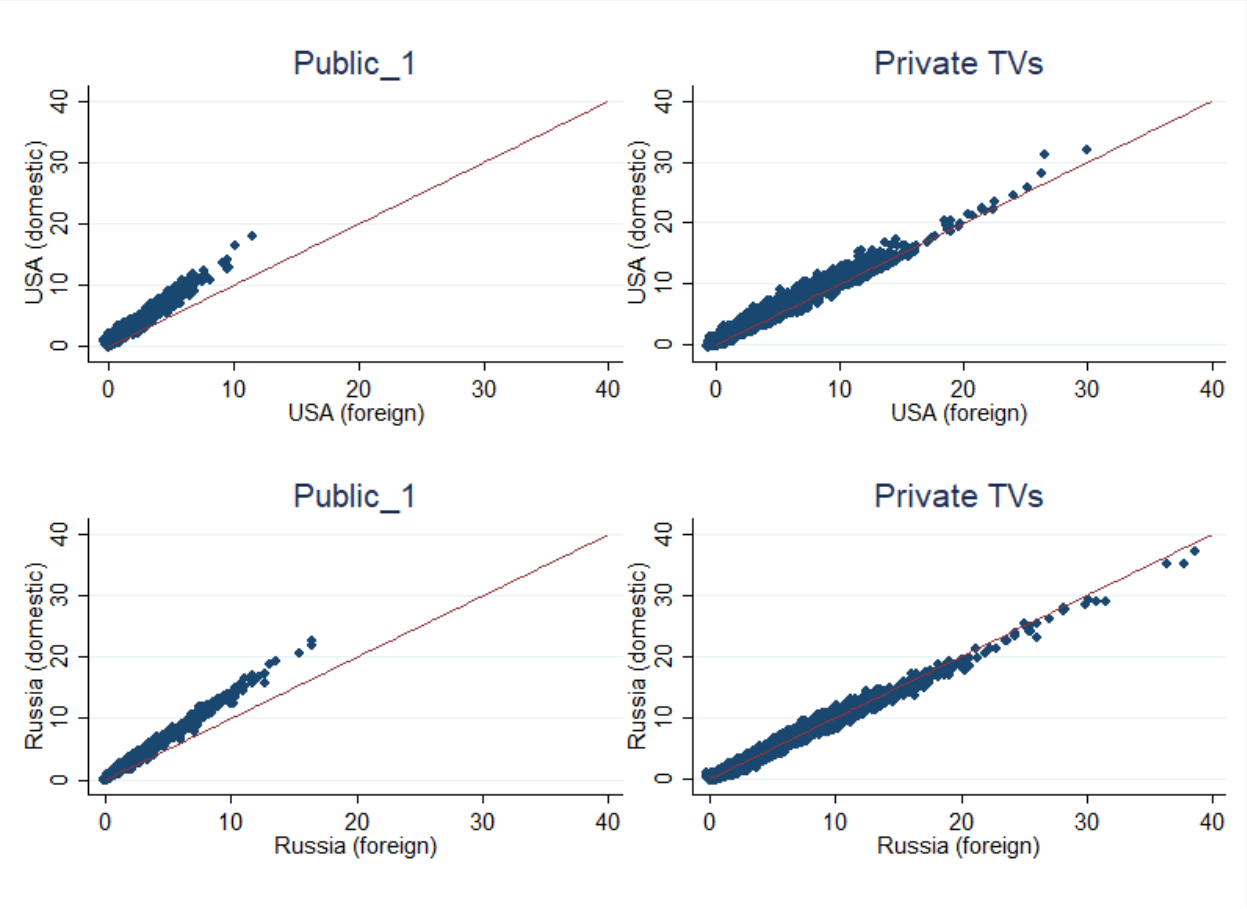
Notes: The figures plot the percentage of TV programs categorized as either the United States topic or the Russia topic for those that are the closest to the official statements (top 7, top 35, and top 70 closest programs): the left figures use the closeness to the Japanese government's statement (top) and the US government's statement (bottom) 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) during President Putin's visit to Japan for the summit. We use machine learning to compute the closeness measures, where the window size is set to 15. See the main text for more details.

Figure VI: Predictability of the machine learning, United States and Russia topics (window=5)



Notes: The figures plot the percentage of TV programs categorized as either the United States topic or the Russia topic for those that are the closest to the official statements (top 7, top 35, and top 70 closest programs): the left figures use the closeness to the Japanese government's statement (top) and the US government's statement (bottom) 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) during President Putin's visit to Japan for the summit. We use machine learning to compute the closeness measures, where the window size is set to 5. See the main text for more details.

Figure VII: Average closeness to the official statements interacted with individuals' viewing time



Notes: These figures plot the closeness to the official statements interacted with individuals' viewing time (AvgExposure in (3)). Each data point represents an individual. The figure shows the results for Public 1 and all private channels. In the top-left panel, the closeness between Public 1's programs and the Japanese government's statement (interacted with duration) (y-axis) is plotted against the closeness between its programs and the US government's statement (interacted with duration) (x-axis), both made when Prime Minister Abe visited Pearl Harbor with President Obama. The top-right panel shows the same, except that closeness is measured for the private channels. Since all five private channels are plotted on the same figure, the number of data points in this panel is five times larger than that in the top-left panel. The bottom panels show the same, except that closeness is measured for the official statements made by the Japanese and Russian governments at the time of the Japan–Russia summit. We use machine learning to compute the closeness measures, where the window size is set to 5, and the vector size is set to 200. See the main text for more details.