PAP for Study 2: South Korea and the US

August 31, 2025

1 Overview

We aim to explore the effect of providing information that different partisan groups are equally knowledgeable in terms of judging true or false about several facts. Specifically, we explore (1) the effect on disbeliefs, (2) the effect on in-group bias in information processing, and (3) the effect on affective polarization. We focus on right-wing party supporters and left-wing party supporters in South Korea and the US. In South Korea, we focus on supporters of the People Power Party and the Democratic Party of Korea. In the US, we focus on supporters of the Republican Party and the Democratic Party. Hereafter, we refer to them as R supporters and L supporters, respectively, according to the conventional partisan labels.

2 Experiment Design

The main part of this experiment contains nine tasks where respondents are asked to judge whether a statement is true or false: $j \in \{1, ..., 9\}$.

There are four types of tasks. First, we have two tasks on factual questions before the treatment. Second, after the treatment, we have three tasks on factual questions and two tasks on conspiracy theory questions with partisan signals. Finally, we have two tasks on factual questions with education group signals.

2.1 Survey Flow

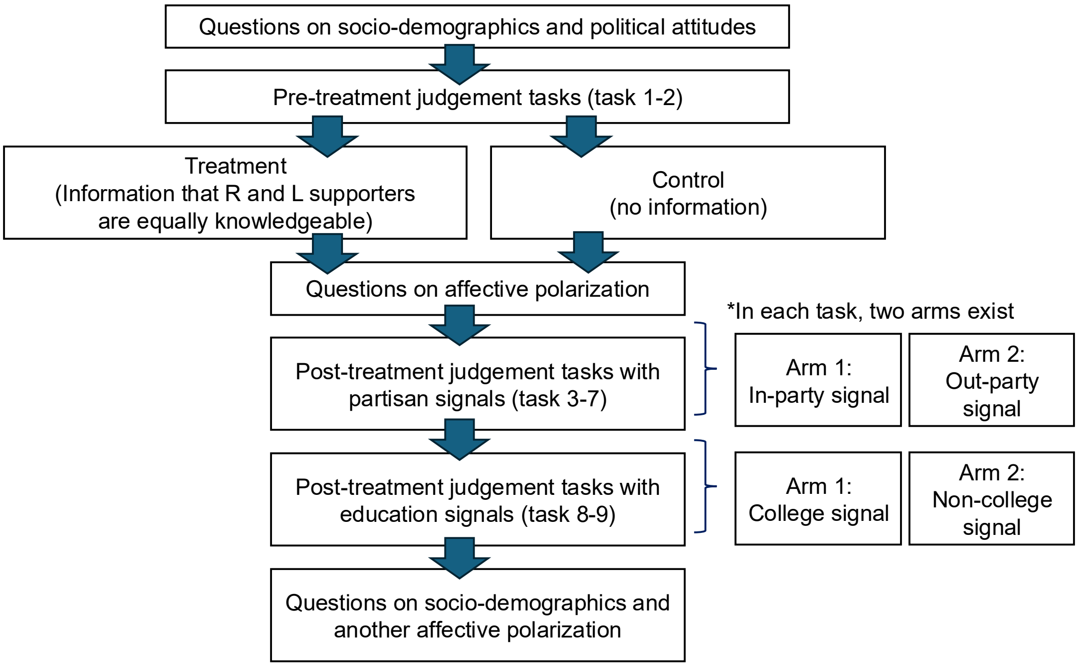
Questions on socio-demographics and political attitudes
Pre-treatment judgement tasks (task 1-2)
Treatment (Information that R and L supporters are equally knowledgeable)
Control (no information)
Questions on affective polarization
*In each task, two arms exist
Post-treatment judgement tasks with partisan signals (task 3-7)
Arm 1: In-party signal
Arm 2: Out-party signal
Post-treatment judgement tasks with education signals (task 8-9)
Arm 1: College signal
Arm 2: Non-college signal
Questions on socio-demographics and another affective polarization

Figure 1: Survey Flow of the Experiment

Socio-demographics, political attitudes, and risk attitudes We first ask about socio-demographics and political attitudes. We ask respondents about gender, age groups, parties they support, and risk attitudes.

Pre-treatment judgement tasks ($j = 1, 2$) After socio-demographic questions, respondents are asked to conduct two true or false judgement tasks on simple facts ($j = 1, 2$). In each task, each respondent is asked to judge true or false of the statement, rate their confidence in their answer, and guess the accuracy rates of R supporters and L supporters.

Treatment Subsequently, half of the respondents will receive information that the difference in the accuracy rates between R supporters and L supporters is less than 5% in all of these tasks. This is the treatment information of our experiment. The control group does not receive any information.

Warmth towards partisan groups We ask respondents how favorable their feelings are toward R supporters and L supporters to measure affective polarization.

Post-treatment judgement task with partisan signals ($j = 3, 4, 5, 6, 7$) Subsequently, respondents are asked to do five additional judgement tasks $j = 3, 4, 5, 6, 7$. Tasks $j = 3, 4, 5$ are about whether a simple fact is true or false. On the other hand, tasks $6, 7$ are about conspiracy questions.

Each task proceeds as follows.

1. Each respondent is asked to judge whether the statement (X) is true or false, rate their confidence in their answer, and guess the accuracy rates of R supporters and L supporters.
2. Each respondent randomly receives one of the following two *signals*: the signal telling them the majority of R supporters' opinion and the signal telling them the majority of L supporters' opinion. The signal is independently drawn across respondents and across tasks.
   - For factual questions, we present the correct answers, which are endorsed by both the majority of R/L supporters
   - For conspiracy theories, we present the position most supported by the majority of R/L supporters
3. Respondents are again asked to judge whether the statement is true and rate their confidence in their answer.

The signal is used to estimate the degree of in-group bias in information processing.

For R (resp. L) supporters, the signal about R (resp. L) supporters' opinions is the in-party signal, whereas the signal about L (resp. R) supporters' opinions is the out-party signal. Thus, each task has two arms—the in-party and out-party signals—with respondents randomly assigned to one.

Post-treatment judgement task with education group signals ($j = 8, 9$) Finally, respondents are asked to do two additional judgement tasks $j = 8, 9$.

Each task proceeds as follows.

1. Each respondent is asked to judge whether the statement (X) is true or false and rate their confidence in their answer.
2. Each respondent randomly receives one of the following two *signals*: the signal telling them the majority of college-graduates' opinion and the signal telling them the majority of non-college-graduates' opinion. The signal is independently drawn across respondents and across tasks.
3. Respondents are again asked to judge whether the statement is true and rate their confidence in their answer.

This signal is used to estimate the degree of bias in information processing towards the college-educated, as we explain later.

Affective polarization Respondents are asked another set of questions on affective polarization.

Socio-demographic questions and risk attitudes We ask another set of socio-demographic questions, such as places of residence and income groups, and risk attitudes.

2.2 Sample Selection

We restrict samples with the following criteria about the baseline level of disbelief.

For each individual i in group $g \in \{R, L\}$ and task j, define the disbelief

$$\text{disbelief}_{i,g(i),j} := p^{g}_{i,g(i),j} - p^{g'}_{i,g(i),j}$$

- $p^{g}_{i,g(i),j}$: estimated accuracy rate towards in-group g
- $p^{g'}_{i,g(i),j}$: estimated accuracy rate towards out-group g'

A larger value means that a respondent estimates the accuracy rate higher for the in-group than for the out-group, implying larger disbeliefs against out-groups.

Then, the pre-treatment disbelief is given by

$$\text{disbelief}^{pre}_{i,g(i)} := \frac{1}{2}\sum_{j=1}^{2} \text{disbelief}_{i,g(i),j}.$$

In the following analysis, we restrict our attention to those with $\text{disbelief}^{pre}_{i,g(i)} > 0.05$. This is because the treatment is expected to reduce disbeliefs only among those with high enough $\text{disbelief}^{pre}_{i,g(i)}$.

3 Analysis Plan

3.1 Treatment Effect on Disbeliefs

The first hypothesis is that the treatment of receiving information about actual accuracy rates for pre-treatment factual questions decreases disbelief in the out-group's knowledge about post-treatment factual questions.

Hypothesis 1: Treatment effect on disbeliefs for factual questions

The post-treatment disbelief regarding factual questions is smaller in the treatment group than in the control group.

Measurement The ex-post disbelief about simple facts is

$$\text{disbelief}^{post,f}_{i,g(i)} := \frac{1}{3}\sum_{j=3}^{5} \text{disbelief}_{i,g(i),j}. \tag{1}$$

Define $T_i = 1$ if individual i is treated.

Specification for Hypothesis 1: We run the following regression:

$$\text{disbelief}^{post,f}_{i,g(i)} = \alpha T_i + const. + \varepsilon_i. \tag{2}$$

Hypothesis 1: We expect the following:

- **Hypothesis 1.** The post-treatment disbelief regarding factual questions is smaller in the treatment group than in the control group. That is, $\widehat{\alpha} < 0$.

3.2 In-group Bias in Information Processing

Because of disbeliefs, people may outweigh the opinion of an in-party member than the opinion of an out-party member in forming their opinion. We call this *in-group bias in information processing.*

Hypothesis 2-1: In-group bias in information processing

Partisans have an in-group bias in information processing for factual questions in the control group.

Hypothesis 2-2: Treatment effects on in-group bias in information processing

Partisans have a smaller in-group bias in information processing for factual questions in the treated group than in the control group.

We use the post-treatment judgment tasks with partisan signals to estimate the degree of such in-group bias in information processing. We focus on the tasks on simple facts ($j = 3, 4, 5$). Analysis for conspiracy theories ($j = 6, 7$) will be conducted as the supplementary analysis.

Measurement Let respondent i's judgment in task j before signals be $J_{i,j,0} \in \{0, 1\}$, where $J_{i,j,0} = 1$ if and only if i's judgment on fact j before the signal is correct. Furthermore, let the estimated accuracy of their own judgment before the signal be $a_{i,j,0} \in [0, 100]$. Then, we define

$$\mu_{i,j,0} = \begin{cases} \frac{a_{i,j,0}}{100} & \text{if } J_{i,j,0} = 1 \\ 1 - \frac{a_{i,j,0}}{100} & \text{if } J_{i,j,0} = 0 \end{cases} \tag{3}$$

Here, $J_{i,j,0}$ is respondent i's *binary opinion* on task j. On the other hand, $\mu_{i,j,0}$ is the *continuous* opinion.

Similarly, let respondent i's judgment on fact j after the signal be $J_{i,j,1}$ and the estimated accuracy of their own judgment after the signal be $a_{i,j,1} \in [0, 100]$. Then, we define

$$\mu_{i,j,1} = \begin{cases} \frac{a_{i,j,1}}{100} & \text{if } J_{i,j,1} = 1 \\ 1 - \frac{a_{i,j,1}}{100} & \text{if } J_{i,j,1} = 0 \end{cases} \tag{4}$$

(J_{ij0}, J_{ij1}) and (μ_{ij0}, μ_{ij1}) serve as measurements of each respondent's binary and continuous opinions before and after signals.

Given these variables, we construct the following two variables for changes in the respondent's opinion. First, let the "dummy update", $y^J_{i,j} \in \{0, 1\}$, where $y^J_{i,j} = 1$ if and only if $J_{i,j,0} = 0$ and $J_{i,j,1} = 1$. This measures if respondents update their beliefs from incorrect to correct answers using only the T/F dichotomy response. Second, let the "continuous update", $y^\mu_{i,j} \in \{0, 1\}$, where $y^\mu_{i,j} = 1$ if and only if $\mu_{i,j,1} > \mu_{i,j,0}$. This measures if respondents update their beliefs from incorrect to correct answers using both the T/F dichotomy response and the continuous responses to the question that asks about the level of confidence in their own T/F responses.

Specification for Hypothesis 2-1: Let s_{ij} be the signal respondent i receives in task j. $s_{ij} = I$ if the signal is about in-party members' opinions and $s_{ij} = O$ if the signal is about out-party members' opinions.

Let $y_{i,j} = \{y^J_{i,j}, y^\mu_{i,j}\}$ be the change in respondent i's opinion on task j before and after the signals. We estimate the following separately for the treated group ($T_i = 1$) and the control group ($T_i = 0$).

$$y_{i,j} = \beta \mathbb{1}\{s_{i,j} = I\} + \eta_j + \varepsilon_{i,j} \tag{5}$$

We denote the estimands of β for the treated group β^T and the control group β^C.

We use both measures $(y^J_{i,j}, y^\mu_{i,j})$ as $y_{i,j}$. In the post-teratment judgement tasks on simple facts, the majority in both political parties correctly give the correct answer, based on our previous survey. Thus, the content of the signal is the same across the two signals. Thus, partisans have an in-group bias in information processing if $\beta > 0$. In other words, β represents the degree of in-group bias in information processing.

We include task fixed effects, η_j, to isolate any unobserved heterogeneity in the propensity of updating beliefs against each task.

Specification for Hypothesis 2-2: For Hypothesis 2-2, we interact the in-group signal dummy with the treatment as follows.

$$y_{i,j} = \beta_1 \mathbb{1}\{s_{i,j} = I\} + \beta_2 T_i + \beta_3 \left(\mathbb{1}\{s_{i,j} = I\} \times T_i\right) + \eta_j + \varepsilon_{i,j} \tag{6}$$

Hypotheses 2-1 and 2-2: We expect the following:

- **Hypothesis 2-1.** Partisans have an in-group bias in information processing for simple facts in the control group. Specifically, $\widehat{\beta} > 0$ in (5) holds in the control group.
- **Hypothesis 2-2.** The treatment reduces in-group bias in information processing for simple facts. That is, $\widehat{\beta_3} < 0$ in (6) holds.

3. Treatment Effect on Affective Polarization

Hypothesis 3: Treatment effect on affective polarization

The treatment decreases the affective polarization.

Measurement We measure respondents' unfavorable feelings toward out-party members and in-party members. The difference between them is:

$$\text{unfav}_{i,g(i)} := \text{fav}^{g}_{i,g(i)} - \text{fav}^{g'}_{i,g(i)},$$

where $\text{fav}^{g}_{i,g(i)}$ is the degree of favorable feelings toward in-party members and $\text{fav}^{g'}_{i,g(i)}$ is the degree of favorable feelings toward out-party members. This is our measurement of affective polarization.

Favorable feelings are measured in two ways. First, we ask respondents to rate positive feelings toward out-party members and in-party members from 0 to 100. We use this as the primary measurement. Second, we ask respondents to answer whether it is (un)comfortable to have a relationship as colleagues/friends/children's spouses. By aggregating them, we construct the second measurement.

Specification for Hypothesis 3: We estimate the following:

$$\text{unfav}_{i,g(i)} = \text{const.} + \gamma T_i + \varepsilon_i. \tag{7}$$

Hypothesis 3: We expect the following:

- **Hypothesis 3.** The treatment reduces affective polarization. That is, $\widehat{\gamma} < 0$.

4 Sample Size

Our main hypothesis is Hypothesis 2-2, and we implement a power analysis based on Hypothesis 2-2.

The following table is the result of our Pilot survey in the US. We collected 466 samples, and 318 samples (68.2%) meet the sample selection criteria discussed in Section 2.2. We have three post-treatment factual tasks, so the number of observations in the regression based on (6) would be 954. We run regressions for $y^J_{i,j}$ and $y^\mu_{i,j}$ separately.

Suppose we set MDE (Minimum Detectable Effect) to be 2/3 of the baseline effect of 0.09, the estimates in both columns. With the power of 0.8 and the significance level of 0.05, we need a sample size of 4,252 for Column (1) and 1,782 for Column (2). Thus, we target 4,252 for both countries.

Table 1: Hypothesis 2-2: Pilot Results from the US

	(1) Change (J)	(2) Change (μ)
In-Group Signal	0.091	0.091
	(0.038)	(0.024)
Treatment	0.023	0.088
	(0.039)	(0.051)
In-Group Signal × Treatment	-0.089	-0.137
	(0.072)	(0.048)
Observations	954	954

Note: The table shows the results of the regressions based on (6) using the data from the pilot survey in the US. Column (1) uses $y^J_{i,j}$ as a dependent variable, while Column (2) uses $y^\mu_{i,j}$. Both columns include task fixed effects. Standard errors are in parentheses and clustered at the individual and task levels.

5 Supplementary Analysis

5.1 Information Processing for Conspiracy Theory Questions

In Hypothesis 2, we exclusively focus on the judgment tasks on simple facts. However, it might be the case that the treatment also affects in-group bias in information processing about conspiracy theories.

Measurement First, we define the ex-post disbelief about conspiracy theories by

$$\text{disbelief}^{post,c}_{i,g(i)} := \frac{1}{2}\sum_{j=6}^{7} \text{disbelief}_{i,g(i),j}. \tag{8}$$

Second, we need to modify our outcome variable $y_{i,j} = \{0, 1\}$. The partisan signals given for conspiracy theory questions presented as tasks $j = 6, 7$ differ between R and L supporters. This is because, for example, the majority of L (resp. R) supporters believe in left-wing (resp. right-wing) conspiracy theory, $j = 6$ (resp. $j = 7$).

Thus, we modify our outcome variable $y_{i,j} = \{0, 1\}$ as follows. Let us denote the type of conspiracy theory tasks $g(j) = L$ for $j = 6$ and $g(j) = R$ for $j = 7$. We define the information updating in conspiracy theory tasks $y^C_{i,j}$ as follows

$$y^C_{i,j} = \begin{cases} y_{i,j} & \text{if } g(i) = g(j) \\ 1 - y_{i,j} & \text{if } g(i) \neq g(j) \end{cases} \tag{9}$$

Specification First, we run the following regression to examine the impact on disbeliefs:

$$\text{disbelief}^{post,c}_{i,g(i)} = \alpha T_i + const. + \varepsilon_i. \tag{10}$$

Second, to examine in-group bias in information processing, we run the following regression:

$$y^C_{i,j} = \beta^I \mathbb{1}\{s_{i,j} = I\} + \beta^O \mathbb{1}\{s_{i,j} = O\} + \eta_j + \varepsilon_{i,j}. \tag{11}$$

- We run this for the treated and control groups separately
- We use both $y^J_{i,j}$ and $y^\mu_{i,j}$ for $y_{i,j}$ as before
- As a robustness check, we also run versions that focus on (a) R supporters for the R conspiracy task and (b) L supporters for the L conspiracy task

Hypotheses First, regarding the effect on disbeliefs, we expect

- **Hypothesis 1-A.** The post-treatment disbelief regarding conspiracy theories is smaller in the treatment group than in the control group.

Second, the majority's opinion is divided across political parties. In particular, the majority of R (resp. L) supporters think that the right-wing conspiracy theory is true (resp. false) but the left-wing conspiracy theory is false (resp. true). Thus, we expect

- **Hypothesis 2-A1.** Signals affect information processing. That is, $\widehat{\beta^I} < 0$ and $\widehat{\beta^O} > 0$ for both the control and the treated groups.
- **Hypothesis 2-A2.** Partisans have an in-group bias in information processing for conspiracy theory in the control group. That is, $\widehat{\beta^I} + \widehat{\beta^O} < 0$ for the control group.
- **Hypothesis 2-A3.** Partisans have a smaller in-group bias in information processing for conspiracy theory in the treated group than in the control group. That is, $|\widehat{\beta^I} + \widehat{\beta^O}|$ for the treated group is smaller than $|\widehat{\beta^I} + \widehat{\beta^O}|$ for the control group.

5.2 Information Processing with Education Signals

To benchmark the size of in-group bias in information processing with partisan signals in the main analysis, we compare it to the bias in information processing when respondents are given signals across different education groups. Specifically, we compare how much respondents update their beliefs based on college graduates' opinions compared to non-college graduates' opinions.

Specification We restrict samples (i,j) to the control group whose pre-signal answers are wrong. We run the following two regressions.

$$y_{i,j} = \tilde{\beta}_1 \mathbb{1}\{s_{i,j} = I\} + const. + \varepsilon_{i,j} \tag{12}$$

$$y_{i,j} = \tilde{\beta}_2 \mathbb{1}\{s_{i,j} = \text{College}\} + const. + \varepsilon_{i,j} \tag{13}$$

By comparing $\widehat{\tilde{\beta}_1}$ and $\widehat{\tilde{\beta}_2}$, we get a sense of the magnitudes of in-group bias in information processing across partisan groups, relative to the bias based on education groups.

5.3 Experimenter demand effect

To ensure that the results are not driven by the experimenter demand effect, we redo the analyses when we exclude the respondents who pander to a hypothesis presented by the experimenter.

Specifically, we exclude those who change their answers to the following question about risk attitudes between the beginning and the end of the survey.

We ask in the survey: "We ask about your attitude towards risk. Suppose that according to the weather forecast, the probability of rain today is 35%. In such a case, do you usually take an umbrella when you go out?" Then, we ask the same question at the end of the survey, but we add "Our hypothesis is that people dislike risks, so they usually take an umbrella" for those who answered "No" at the beginning of the survey. Similarly, we add "Our hypothesis is that people like risks, so they usually do not take an umbrella" for those who answered "Yes" at the beginning of the survey.

If the answers differ between the beginning of and the end of the survey, it would be because a respondent panders to the hypothesis presented by the experimenter. Thus, such respondents are subject to the experimenter demand effect.

5.4 Heterogeniety Analysis

The degree of in-group bias and the treatment effect may vary depending on respondents' characteristics, such as age, gender, and partisanship. We will investigate whether such heterogeneity exists.