

# Distrusting the Out-Party: Partisan Disbelief and Biased Information Processing\*

Shinnosuke Kikuchi  
UCSD

Daiki Kishishita  
Hitotsubashi

Yesola Kweon  
SKKU

Yuko Kasuya  
Keio

January 29, 2026

## Abstract

This paper introduces and tests a new concept—*partisan disbelief in knowledge*—the tendency for partisans to believe that their in-group is more knowledgeable than the opposing party, even about basic non-partisan facts. Using large-scale surveys and experiments in South Korea and the United States, we document that partisans perceive their in-group’s accuracy rate in judging non-partisan facts to exceed that of the out-group by about 15 percentage points. This partisan disbelief distorts information processing: when identical information is attributed to out-group sources, individuals are less likely to update their opinions, revealing an in-group bias that extends beyond explicitly partisan issues. Providing corrective evidence that both sides are equally knowledgeable reduces partisan disbelief and weakens this in-group bias. Together, these results show that polarization extends beyond politics to perceptions of competence, identifying a novel cognitive mechanism through which social identity undermines accurate information processing and mutual understanding in societies.

**Keywords:** Misperception; Partisan bias; Information processing; Polarization

**JEL classification:** D72, D91

---

\*We thank Daron Acemoglu, Toru Kitagawa, Ro’ee Levy, Hirofumi Miwa, Shoko Omori, and Hitoshi Shigeoka for their helpful comments. We thank Gion Toyokawa for excellent research assistance. We also thank seminar participants at the JSQPS Winter Meeting, the Japanese Public Choice Society Meeting, and UCSD. This study was financially supported by the JSPS Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research Grant Number JPJS00123811919.

# 1 Introduction

Political polarization has become one of the most pressing challenges for contemporary democracies. According to the Pew Research Center, a median of 65% of respondents across 19 countries report that their societies are experiencing strong partisan conflict (Silver et al., 2022). In highly polarized environments, party identification functions as a salient social identity: individuals regard co-partisans as in-group members and opposing partisans as out-group members (e.g., Green et al., 2002; Huddy et al., 2015). Such identity divisions can distort how people process information and interact with one another in everyday life.

A large body of research in economics, political science, and psychology shows that social identity fosters systematic misperceptions about out-group members, including in partisan contexts (see Bursztyn and Yang, 2022). People often misjudge the diversity of opinions within the opposing party (Dias et al., 2025) and overstate the degree of ideological conflict (Ahler, 2014). Yet the literature has paid little attention to a more fundamental dimension of misperception—how partisans perceive others' *knowledge*.

This paper studies a novel form of bias which we term *partisan disbelief in knowledge*. We posit that partisans may believe that supporters of their own party are more knowledgeable than supporters of the opposing party, even about basic *non-partisan* facts (e.g., whether the iPhone was invented before 2000). If present, this belief represents a new channel through which polarization distorts cognition and communication.

Partisan disbelief matters for two reasons. First, it can generate an in-group bias in information processing. When an individual encounters information attributed to an out-group source, partisan disbelief may lead them to discount it as unreliable—hindering information aggregation in everyday interactions and on social media, even in the absence of network homophily (Golub and Jackson, 2012). This mechanism differs from well-studied *partisan selective exposure* (e.g., Peterson et al., 2021; Faia et al., 2024; Chopra et al., 2024), which arises because individuals view out-group information as ideologically biased. In contrast, partisan disbelief implies distrust even for non-partisan information, producing spillovers of polarization beyond the political domain. As a result, partisan disbelief not only shapes political attitudes but also weakens the shared fact-based communications and collective reasoning.

Second, partisan disbelief may amplify affective polarization. Contemporary political conflict is not only ideological but also emotional: many citizens dislike and distrust the out-party members (Iyengar and Westwood, 2015; Iyengar et al., 2019; Boxell et al., 2024). Affective polarization may even undermine democratic norms, although the evidence is mixed (Druckman and Levendusky, 2019; Voelkel et al., 2023; Cox et al., 2025), and it distorts social interactions even in non-political domains (e.g., Fowler and Kam, 2007; Huber and Malhotra, 2017; Shafranek, 2021; Mill and Morgan, 2022; Dimant, 2024). If people view out-groups as less knowledgeable, such perceptions can intensify distrust and contempt, reinforcing affective polarization.

Motivated by these considerations, we empirically examine the existence and consequences of partisan disbelief. We address three questions: (1) Does partisan disbelief exist in polarized

societies? (2) Does it create an in-group bias in information processing, and can corrections reduce that bias? (3) Does reducing partisan disbelief mitigate affective polarization?

To answer these questions, we conducted two online studies—a baseline survey and a survey experiment—in two highly polarized democracies, South Korea and the United States. Pew surveys indicate that 90% of South Koreans and 88% of Americans perceive intense partisan conflict (Silver et al., 2022). Both countries have clear two-party systems, allowing clean identification of in- and out-groups, yet they differ substantially in culture, institutions, and democratic history, which enhances the external validity of our findings.

In the baseline survey (approximately 1,500 respondents per country), participants evaluated eight true-or-false statements on non-partisan facts (both political and non-political). For each statement, respondents judged whether it was true or false and estimated what share of supporters in each political group would answer correctly. Comparing perceived accuracy rates across groups allows us to measure partisan disbelief—the tendency to think that in-group members are better informed than out-groups, even on non-political facts.<sup>1</sup>

We find that partisans perceive the accuracy rate of in-party members in these judgment tasks to be higher than that of out-party members by approximately 15 percentage points for both partisans in the US and South Korea. Given that the actual accuracy rates differ across partisan groups by less than 5 percentage points, this misperception is both substantial and striking.

Building on these findings, we implemented a large-scale survey experiment (approximately 4,200 respondents per country) to test the mechanisms linking partisan disbelief to information processing and affective polarization. Participants first answered factual questions and estimated partisan accuracy rates, enabling us to measure each respondent’s baseline disbelief. We then randomly provided half of the respondents with corrective information showing that, in our baseline data, supporters of both parties were almost equally accurate. Next, all participants received “signals” reporting the majority judgment of either in-group or out-group supporters on additional factual statements. Because the content of the signals was identical across groups, any differential updating reveals in-group bias in information processing.

The experiment yields three main findings. First, corrective information significantly reduces partisan disbelief, demonstrating that the bias is malleable. Second, we document clear in-group bias in information processing—stronger in South Korea than in the United States—and show that reducing disbelief weakens this bias. Third, providing corrective information lowers one measure of affective polarization, though the effect fades by the end of the survey, indicating that such interventions do not have a large impact.

Together, these results reveal that partisans not only distrust the opposing side’s motives but also *underestimate its knowledge*. This misperception distorts information processing even on apolitical issues.

---

<sup>1</sup>In the current draft, we rely on results from experiments without incentives. We are currently in the process of preparing for experiments with incentives, where respondents will get more payments if their estimations on the accuracy for each group are closer.

**Related literature** This study is related to three strands of the literature: (i) misperceptions about others, (ii) in-group bias in information selection and processing, and (iii) interventions to reduce affective polarization.

First, numerous studies in economics, political science, and psychology document that people hold systematic misperceptions about others, which are particularly pronounced when directed at out-groups (see [Bursztyn and Yang \(2022\)](#) for a review).<sup>2</sup> With the rise of affective polarization, partisanship has become a central dividing line between in- and out-groups ([Iyengar et al., 2019](#)), making misperceptions about out-partisans especially likely. Prior work shows that people misperceive the demographic composition of parties ([Ahler and Sood, 2018](#)), the diversity of opinions within the opposing party ([Dias et al., 2025](#)), the extent of ideological polarization ([Ahler, 2014](#)), the degree of affective polarization ([Druckman et al., 2022](#)), and the likely behavior of out-groups in strategic settings ([Dimant, 2024](#)). Relatedly, work in economics shows that misperceptions shape redistribution preferences (e.g., [Alesina et al., 2018](#)).<sup>3</sup> Our contribution is to identify and test a new type of misperception—*partisan disbelief in knowledge*: the tendency to think co-partisans are more knowledgeable than out-partisans even about apolitical facts. Unlike previously studied misperceptions, this bias is unrelated to partisan *content* and concerns the perceived *competence* of the source.

Second, we contribute to the literature on in-group bias in information selection and processing. Citizens today differ not only in ideological positions but also in perceptions of factual reality ([Alesina et al., 2020](#)). Many studies document partisan selective exposure (in-group biased *selection*) (e.g., [Peterson et al., 2021](#); [Faia et al., 2024](#); [Chopra et al., 2024](#); [Dimant et al., 2024](#)), yet curbing selection alone has limited effects on opinions ([Levy, 2021](#)), pointing to in-group bias during *processing*, which is less studied.<sup>4</sup> The literature documents systematic deviations from Bayesian updating, such as motivated reasoning and confirmation bias, in political settings (e.g., [Lord et al., 1979](#); [Taber and Lodge, 2006](#); [Taber et al., 2009](#); [Thaler, 2024](#)) as well as in apolitical settings (e.g., [Fryer Jr et al., 2019](#); [Zimmermann, 2020](#); [De Filippis et al., 2022](#); [Angrisani et al., 2021](#)). However, they are conditional on information content rather than on source identity.<sup>5</sup> We add two points. First, we isolate an in-group processing bias that operates through *distrust in others' knowledge*: identical signals are discounted when attributed to out-partisans. Second, we show that this mechanism arises even on non-partisan facts, beyond the scope of selective exposure or content-driven motivated reasoning. Closely related are [Zhang and Rand \(2023\)](#), who document disbelief in others' ability to detect fake news and a processing bias in the U.S., and [Moorthy](#)

---

<sup>2</sup>Another implication of social identity is that altruism may be extended only to in-group members. The degree to which individuals exhibit the same level of altruism toward strangers as toward in-group members has been measured across many countries ([Enke et al., 2023](#); [Cappelen et al., 2025](#)).

<sup>3</sup>They document misperceptions about intergenerational inequality. See also [Alesina et al. \(2020\)](#).

<sup>4</sup>[Fang et al. \(2025\)](#) find that conversations with contrary-minded individuals did not lead to convergence in political views.

<sup>5</sup>[Hill \(2017\)](#) find symmetric learning when information comes from computers (see also [Moorthy \(2025\)](#)), underscoring source identity. [Kashner and Stalinski \(2024\)](#) show that the order between partisan and non-partisan information matters even with identical content. By contrast, [Faia et al. \(2024\)](#) find that revealing the news source affects *selection* but not *processing* once read.

(2025), who study disbelief in others' ability to apply Bayes' rule in India. We differ in the type of disbelief (perceived general knowledge rather than fake-news detection or Bayesian application), in the non-partisan domain we study, and by showing that *correcting* disbelief attenuates in-group processing bias.

Lastly, we contribute to research on mitigating affective polarization. Existing interventions include correcting misperceptions about out-party composition (e.g., Ahler and Sood, 2018), addressing misperceived motives (e.g., Lees and Cikara, 2020), reducing partisan identity salience (e.g., Levendusky, 2018), highlighting cross-party warmth among leaders (e.g., Huddy and Yair, 2021), and increasing intergroup contact (e.g., Whitt et al., 2021). Recent work advances comparative measurement and causal identification of affective polarization (Gidron et al., 2020, 2023), while new measures refine how polarizing everyday interactions are (Hudde et al., 2024). We add a complementary lever: providing corrective evidence that both sides are similarly knowledgeable reduces partisan disbelief and weakens in-group processing bias; it also temporarily lowers affective polarization. In short, we identify a novel, competence-based channel linking social identity to mutual understanding and show how light-touch corrections can partially mitigate its consequences.

The remainder of the paper is organized as follows. Section 2 describes the design and results of the baseline survey, demonstrating the prevalence of partisan disbelief. Section 3 outlines the experimental design, Section 4 presents the empirical hypotheses, and Section 5 presents the experimental results. Section 6 concludes.

## 2 Prevalence of Partisan Disbelief: Baseline Survey

The first objective of this study is to examine whether partisan disbelief in political knowledge—defined as the misperception that out-group members possess lower knowledge than in-group members—exists in two highly polarized countries: South Korea and the United States.

### 2.1 Case Selection: South Korea and the United States

South Korea provides an instructive case of severe yet distinctive partisan polarization. Silver et al. (2022) report that nearly 90% of South Koreans perceive strong or very strong partisan conflict—one of the highest shares among advanced democracies. The country's two-party alignment between the conservative People's Power Party and the liberal Democratic Party of Korea produces a clear division of political identity, making it easy to define in- and out-groups. Polarization in South Korea has intensified over time as corruption scandals, rapid social change, and generational cleavages over gender and inequality have deepened distrust between partisans (Cheong and Haggard, 2023). These tensions culminated in the 2024 martial-law crisis. Because South Korea's democracy is relatively young and its partisan alignments have evolved rapidly

since democratization in 1987,<sup>6</sup> South Korea offers a useful context for examining how partisanship affects and cognitive bias can emerge in a newer democracy with high political engagement and low cross-party trust.

The United States, by contrast, represents a long-established democracy where polarization has become a defining feature of public life. Approximately 88% of Americans perceive intense partisan conflict, and affective polarization—defined as dislike and distrust of the opposing party—has increased sharply over the past four decades (Silver et al., 2022). The U.S. two-party system provides a clear and stable mapping of political identity, yet its institutional, cultural, and media environment differs sharply from South Korea’s. Studying the United States, therefore, allows us to benchmark the mechanisms of partisan disbelief in a mature democracy with entrenched ideological sorting and extensive exposure to partisan media.

Taken together, these two countries combine analytical clarity in identifying partisanship with contrasting historical and institutional backgrounds, enhancing the external validity of our findings on how polarization shapes perceptions of others’ knowledge.

In what follows, we refer to the People’s Power Party in South Korea and the Republican Party in the United States as *party R*, and the Democratic Party of Korea in South Korea and the Democratic Party in the United States as *party L*.

## 2.2 Survey Design

To investigate whether partisan disbelief in political knowledge exists in the two countries, we conducted a Qualtrics-based online survey in South Korea from May 26 to June 4,<sup>7</sup> and in the U.S. from August 12 to August 15, 2025. Respondents were recruited via an established online panel (PureSpectrum). We employed quota sampling on gender, age, and region of residence to ensure representativeness. Participants received a participation fee from the survey firm. The study was pre-registered on OSF (SK: osf.io/6dyct, U.S.: osf.io/tv52h), and the survey was approved by the Institutional Review Board at Keio University. To ensure respondent quality, we implemented a directed-response (satisficing) check at the start of the survey. Respondents were instructed to demonstrate attention by choosing “I have a question” from the response set (“I understand,” “I do not understand,” “I have a question”). Those who selected any other option were screened out and prevented from continuing. In total, 1,498 respondents completed the survey in South Korea and 1,597 in the United States.

**Key variables of partisan disbelief in knowledge:** To measure partisan disbelief in knowledge, respondents answered 8 true-or-false questions on non-partisan factual statements (both political and non-political). The list of questions is provided in Table 1.

<sup>6</sup>In contrast to the U.S., the party system in South Korea is unstable as seen in the frequent changes in the party names. However, the two streams of political camp, conservative and liberal, are relatively well-defined (Cheong and Haggard, 2023).

<sup>7</sup>The survey was conducted during the presidential election.

For each statement, respondents indicated whether it was true or false and rated their confidence on a 0–100 scale. They were then asked to estimate the percentage of individuals in each of three groups—supporters of party *R*, supporters of party *L*, and non-partisans—who would correctly identify the statement as true or false. These responses form the basis for our measure of partisan disbelief in knowledge.

The survey also included four additional questions that were similar to those described above, except that the statements concerned conspiracy theories—both right-wing and left-wing. These questions were included to examine partisan disbelief regarding partisan facts. The analysis of partisan disbelief in knowledge related to conspiracy theories is provided in the Appendix B.

Note that we did not provide monetary incentives to elicit precise estimates of the accuracy rates for each political group, although doing so would have been feasible. When individuals form beliefs, two considerations are at play: forming accurate beliefs is beneficial in terms of material payoff, but it may reduce psychological payoff by limiting their ability to continue believing what they wish to believe (Little, 2019). As a result, individuals engage in motivated reasoning (Lord et al., 1979; Taber and Lodge, 2006; Taber et al., 2009; Fryer Jr et al., 2019; Zimmermann, 2020; Thaler, 2024). Providing monetary incentives increases the relative weight of material payoff. Hence, the accuracy rates elicited under monetary incentives would likely diverge from the accuracy rates individuals hold in daily life. This concern dominates the advantage of monetary incentives in eliciting partisan beliefs in highly polarized societies. For this reason, we chose not to provide such incentives.

Table 1: List of True-or-False Questions

	South Korea	United States
Non-partisan facts	Political The term of the National Assembly is 2 years. (F) To revise the Constitution, a majority of votes in a national referendum is required. (T) The country's nominal GDP growth rate in the last year was lower than 5%. (T) For every 100 people of working age (15–64), there are 40 people aged 65 or older. (T)	The term of office in the Senate is 4 years. (F) To revise the Constitution, approval from more than three-fourths of the state legislatures is required. (T) The country's nominal GDP growth rate in the last year was lower than 7%. (T) For every 100 people of working age (15–64), there are 40 people aged 65 or older. (F)
	Non-political New Zealand is a country located in the Middle East. (F) The iPhone was invented before 2000. (F) By law, you must be at least 19 years old to drink alcohol. (T) The highest mountain in the country is Hallasan. (T)	New Zealand is a country located in the Middle East. (F) The iPhone was invented before 2000. (F) Alaska is the largest state in the United States. (T) The highest mountain in the United States is Mt. McKinley (Denali). (T)

**Individual-level measurement on polarization:** In addition to measuring partisan disbelief, we also assessed individual-level polarization in order to examine its correlation with partisan disbelief.

First, we measured self-reported affective polarization using two approaches.<sup>8</sup> The first asked respondents: “On a scale from 0 to 100, where 0 means very cold or unfavorable feelings and 100 means very warm or favorable feelings, how warmly do you feel toward supporters of each political party?” The second question asked respondents: “How would you feel about being in the following types of relationships with supporters of each political party: (i) colleagues at work, (ii) close friends, and (iii) your own or your child’s spouse?”

Second, we measured meta-perceived level of affective polarization by asking “On a scale from 0 to 100, where 0 means very cold or unfavorable feelings and 100 means very warm or favorable feelings, how warmly do you think party  $L$  (resp.  $R$ ) supporters feel toward party  $R$  (resp.  $L$  supporters?)”

Lastly, we measured ideological polarization and meta-perceived ideological polarization. Respondents were asked: “On a scale from 0 to 10, where 0 means ‘extremely liberal’ and 10 means ‘extremely conservative,’ where would you place yourself and the average supporter of each political party?” From this, we derive measures of ideological extremity and meta-perceived ideological distance between the two parties.

### 2.3 Summary Statistics

Table 2 presents the summary statistics from the baseline survey. Several observations emerge. First, each political group differs in its demographic composition; for example, in South Korea, the share of respondents aged 50 and above is higher among supporters of the People’s Power Party. Second, in both countries, the average accuracy rates of party  $R$  supporters and party  $L$  supporters do not differ substantially.<sup>9</sup> Third, polarization is pronounced in both countries, in terms of both affective and ideological polarization.

### 2.4 First Look at Partisan Disbelief

With the rise of affective polarization, partisanship has become a central dividing line between in-group and out-group members (Iyengar et al., 2019), making misperceptions about out-party members particularly likely. We expect that partisans believe supporters of the opposing party are less knowledgeable about non-partisan facts than themselves.

<sup>8</sup>There are four measurements for affective polarization frequently used in the literature (Druckman and Levendusky, 2019): (i) feeling thermometers, (ii) ask respondents to rate partisans when it comes to being hypocritical, selfish, honest, or generous, (iii) how much one can trust the parties, and (iv) social distance measures that ask about individuals’ comfort in having their child marry someone from another party or having a friend from the other party. Our first measurement corresponds to (i), while our second measurement corresponds to (iv). Druckman and Levendusky (2019) find that all four measures are correlated, the former three measures document particularly high correlation, but (iv) is less strongly correlated with (i)-(iii). Therefore, we use (i) and (iv).

<sup>9</sup>In the United States, non-partisans’ average accuracy rate is 9 percentage points lower than that of Republican Party supporters.

Table 2: Summary Statistics: Baseline Survey

	SK RP	SK LP	SK NP	US RP	US LP	US NP
<b>Demographics</b>						
Female Ratio	0.41	0.49	0.60	0.47	0.59	0.55
College-educated Ratio	0.83	0.77	0.76	0.52	0.54	0.38
Age (50+) Ratio	0.54	0.42	0.36	0.54	0.48	0.38
<b>Judgements</b>						
Average Accuracy Rate	0.74	0.74	0.73	0.67	0.63	0.58
Average Confidence	0.81	0.78	0.71	0.74	0.71	0.65
<b>Affective Polarization (0 to 1)</b>						
Warmth toward RP	0.75	0.21	0.39	0.82	0.29	0.46
Warmth toward LP	0.24	0.76	0.41	0.37	0.80	0.51
Comfortableness with RP at Work	0.78	0.49	0.56	0.81	0.68	0.77
Comfortableness with RP as Friend	0.78	0.46	0.54	0.82	0.66	0.77
Comfortableness with RP as Child	0.78	0.43	0.53	0.82	0.63	0.77
Comfortableness with LP at Work	0.47	0.78	0.57	0.68	0.88	0.79
Comfortableness with LP as Friend	0.45	0.79	0.56	0.69	0.88	0.79
Comfortableness with LP as Child	0.43	0.80	0.55	0.67	0.88	0.80
<b>Ideological Polarization (0 to 1)</b>						
Conservatism of Self	0.71	0.39	0.53	0.76	0.39	0.52
Extremity of Self (rel. to Center)	0.47	0.35	0.17	0.59	0.52	0.36
Conservatism of RP	0.78	0.73	0.64	0.79	0.60	0.53
Conservatism of LP	0.23	0.38	0.38	0.25	0.46	0.41
Observations	343	596	458	617	596	384

*Note:* This table shows the summary statistics for the baseline surveys in South Korea and the United States. We report the averages for each country and each partisan group. RP denotes right-wing party supporters, LP denotes left-wing party supporters, and NP denotes non-partisans.

To formally explore this hypothesis, let  $i$  denote individuals and  $g \in \{R, L, N\}$  denote one of three political groups: supporters of party  $R$  ( $R$ ), supporters of party  $L$  ( $L$ ), and non-partisans ( $N$ ). Let  $g(i)$  be the party individual  $i$  supports. Let also  $j$  index the judgment tasks on non-partisan statements  $j \in \{1, \dots, 8\}$ . We then define  $p_{i,j}^t$  as individual  $i$ 's estimated accuracy rate towards target group  $t$ .<sup>10</sup>

**Results** Figure 1 shows the distributions of the averages of  $p_{i,j}^t$  across different factual question tasks ( $j = 1, 2, \dots, 8$ ) for each perceiver and target group in South Korea, together with the actual accuracy rates. As confirmed in Table 2, the actual accuracy rates are similar across political groups. Nevertheless, the median supporter of party  $R$  perceives the in-party accuracy rate to be higher than 75%, whereas supporters of party  $L$  perceive it to be around 62.5%. A similar pattern holds for supporters of party  $L$ . In contrast, non-partisans perceive the accuracy rates of party  $R$  and party  $L$  supporters to be roughly the same. This represents partisan disbelief in knowledge. Specifically, it captures how each perceiver changes their estimation of the accuracy rate depending on the target. Therefore, we refer to this as *target-based partisan disbelief*.

Another way to capture partisan disbelief is to examine how the estimated accuracy rate of a particular political group differs depending on the perceiver's identity. For example, regarding the accuracy rate of party  $R$  supporters, the median supporter of party  $R$  predicts it to be higher than 75%, whereas the median supporter of party  $L$  predicts it to be lower than 62.5%. A similar pattern holds when the target is party  $L$  supporters. However, when the target group is non-partisans, such bias is not observed. We refer to this as *perceiver-based partisan disbelief*.

Figure 2 shows the distributions of the average of  $p_{i,j}^t$  across different factual question tasks ( $j = 1, 2, \dots, 8$ ) for each perceiver and target group in the U.S., along with the actual accuracy rates. Although the exact numbers differ from those in South Korea, we observe both target-based and perceiver-based partisan disbelief in the U.S..

So far, we have graphically examined the partisan disbelief in the two countries. In the next two sections, we formalize and empirically test the existence of partisan disbelief.

## 2.5 Target-Based Partisan Disbelief

We first formally define target-based partisan disbelief. Citizens have target-based partisan disbelief if the following two conditions are satisfied:

- (a) Supporters of each party believe that members of their own party are more knowledgeable than members of the opposing party. Formally, for each perceiver group  $g(i) \in \{R, L\}$  separately, we estimate

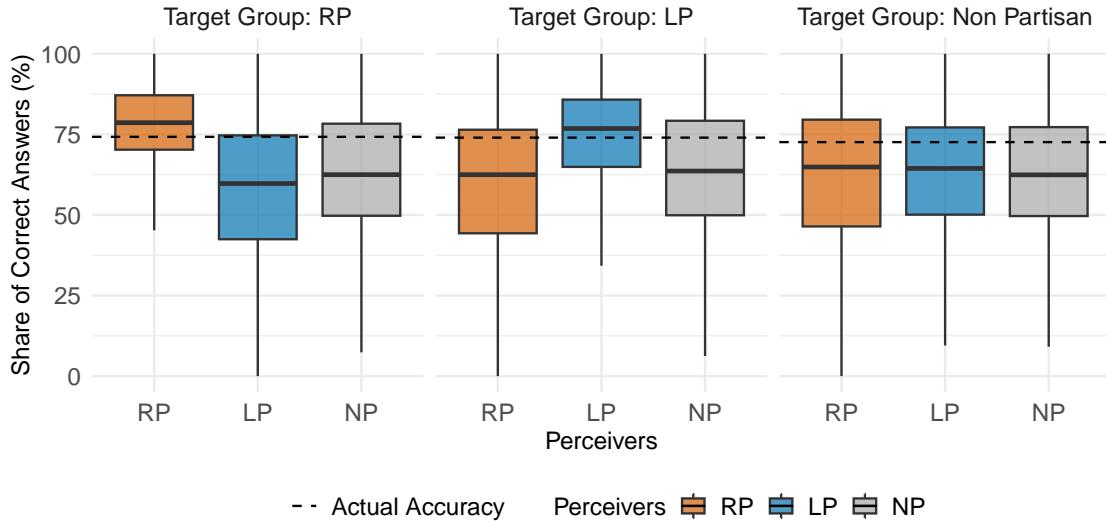
$$p_{i,j}^t = \beta_1 \mathbb{1}\{t = i\} + \mu_i + \mu_j + \varepsilon_{i,j}^t, \quad (1)$$

where  $\mu_i$  and  $\mu_j$  are individual- and issue-fixed effects. We only use the data for targets being either  $R$  or  $L$ . We expect  $\beta_1 > 0$ .

---

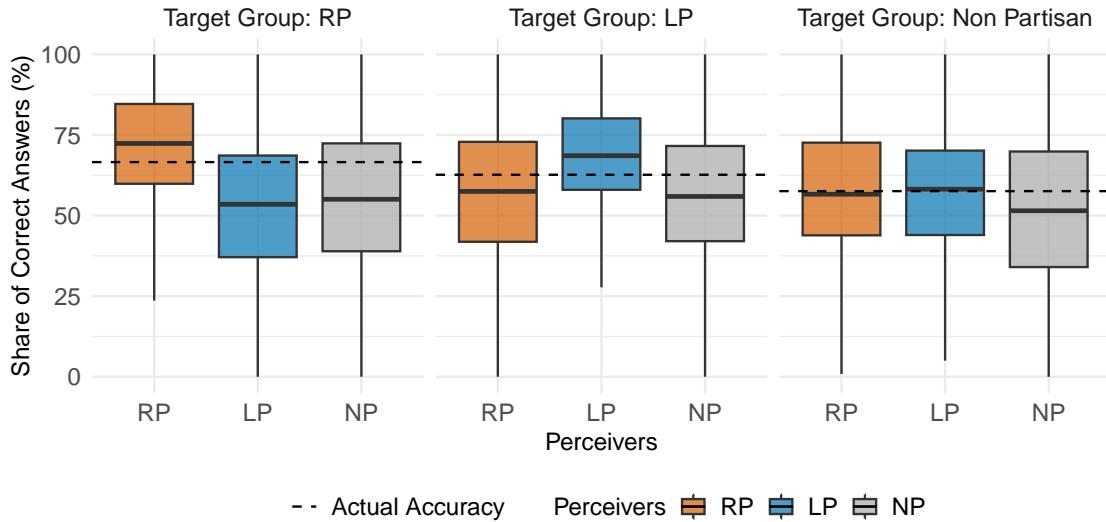
<sup>10</sup>The question was originally asked on a 0–100 scale, which we normalized to a 0–1 scale.

Figure 1: Partisan Disbelief in South Korea: All Factual Questions



Notes: The figure plots the distributions of perceived accuracy, measured as the average of  $p_{i,j}^t$  across all eight factual questions ( $j = 1, \dots, 8$ ), for each combination of perceiver and target group in South Korea. Each box represents the interquartile range (25th–75th percentile), with the median indicated by a solid black line. The dashed horizontal line denotes the actual average accuracy of each target group.

Figure 2: Partisan Disbelief in the United States: All Factual Questions



Notes: The figure plots the distributions of perceived accuracy, measured as the average of  $p_{i,j}^t$  across all eight factual questions ( $j = 1, \dots, 8$ ), for each combination of perceiver and target group in the United States. Each box represents the interquartile range (25th–75th percentile), with the median indicated by a solid black line. The dashed horizontal line denotes the actual average accuracy of each target group.

- (b) By contrast, non-partisans are expected to view the two parties' supporters as equally knowledgeable. Formally, for individuals with  $g(i) = N$ , we estimate

$$p_{i,j}^t = \beta_2 \mathbb{1}\{t = R\} + \mu_i + \mu_j + \varepsilon_{i,j}^t. \quad (2)$$

We only use the data for targets being either  $R$  or  $L$ . We expect  $\beta_2 = 0$ , which indicates no perceived partisan difference in knowledge among non-partisans.

Target-based partisan disbelief captures within-perceiver in-group favoritism.<sup>11</sup>

**Results** Table 3 presents the results of estimating equations (1) and (2) for both countries. Each column reports estimates separately by country and by the perceiver's partisan group: right-wing party supporters (RP), left-wing party supporters (LP), and non-partisans (NP). The rows labeled "Target = RP" and "Target = LP" report the estimated perceived knowledge gap between RP and LP targets, evaluated from the perspective of the perceiver group in that column (with standard errors in parentheses). In each column, the omitted target category serves as the reference group, so the reported coefficient can be read as the perceived difference in knowledge between the indicated target group and the omitted target group.

Column (1) focuses on RP perceivers in South Korea. On average, they rate the accuracy of party supporters at 65.3%. Relative to that baseline, the coefficient in the "Target = RP" row is 0.174 (s.e. 0.012), implying that RP perceivers rate RP targets as more knowledgeable than LP targets by 17.4 percentage points. Column (2) reports the analogous estimate for LP perceivers in South Korea. Their mean perceived accuracy is 64.4%, and LP perceivers rate LP targets as more knowledgeable than RP targets by 15.1 percentage points (0.151, s.e. 0.007).

Columns (4) and (5) show the same pattern in the United States: RP perceivers have a mean estimate of 61.2% and assign RP targets a 15.8 percentage point advantage over LP targets (0.158, s.e. 0.009), while LP perceivers have a mean estimate of 58.8% and assign LP targets a 14.9 percentage point advantage (0.149, s.e. 0.009). These magnitudes are comparable across the two countries, and the associated standard errors are small relative to the point estimates, indicating that the perceived in-party advantage is precisely estimated in each partisan subsample.

In contrast, the NP columns show little perceived difference across party targets. In South Korea (column (3)), the "Target = RP" estimate is 0.005 (s.e. 0.007), and in the United States (column (6)) it is 0.014 (s.e. 0.010), both close to zero in magnitude. Consistent with this, the mean outcomes for non-partisans are lower—60.4% in South Korea and 53.4% in the United States—indicating that non-partisans, on average, rate targets' accuracy more conservatively, while not differentiating sharply between RP and LP targets.

Overall, the table shows a clear and symmetric partisan disbelief—RP perceivers favor RP targets and LP perceivers favor LP targets—while non-partisans rate the two partisan groups as

---

<sup>11</sup>In Appendix A, we show that the results are the same if we use another definition of partisan disbelief—perceiver-based partisan disbelief.

being almost equally knowledgeable. These results confirm the existence of target-based partisan belief in both countries.

Table 3: Target-based Partisan Disbelief: Baseline Survey

Country Perceiver	SK RP (1)	SK LP (2)	SK NP (3)	US RP (4)	US LP (5)	US NP (6)
Target = RP	0.174 (0.012)		0.005 (0.007)	0.158 (0.009)		0.014 (0.010)
Target = LP		0.151 (0.007)			0.149 (0.009)	
Mean of Outcome	0.653	0.644	0.604	0.612	0.588	0.534
Observations	8232	14304	10992	14808	14304	9216

Note: This table reports the results of estimating equations (1) and (2) that test for target-based partisan disbelief in South Korea (SK) and the United States (U.S.). Columns (1)–(3) use the SK sample and columns (4)–(6) use the U.S. sample. Within each country, columns correspond to the perceiver’s partisan group: supporters of the right party (RP), supporters of the left party (LP), and non-partisans (NP). The dependent variable  $p_{i,j}^t$  is perceiver  $i$ ’s assessment of the accuracy (probability correct) of target group  $t \in \{\text{RP}, \text{LP}\}$  on issue  $j$ . The rows “Target = RP” and “Target = LP” indicate which target group’s perceived accuracy is being compared to the omitted target group in that column. All regressions include individual fixed effects and issue fixed effects. Standard errors clustered at the individual level are shown in parentheses.

## 2.6 Correlation with Ideological and Affective Polarization

In addition, we expect that partisan disbelief is correlated with individual-level polarization. While its correlation with affective polarization, emotional and identity-based animosity, is straightforward, additional explanation is warranted for why we also expect a correlation with ideological polarization, which is partisan difference in issue positions and ideological values. First, as partisans increasingly rely on in-group sources and reject out-group expertise due to partisan disbelief, they accumulate different factual premises about the same political or social issues. This asymmetry in belief updating creates self-reinforcing ideological consistency within each group. Second, when individuals observe that the opinions of out-group members diverge from those of in-group members, they may attribute this divergence to a perceived lack of knowledge among out-group members rather than to fundamental differences in preferences.<sup>12</sup> Therefore, ideological polarization may in turn reinforce partisan disbelief. Therefore, we expect partisan disbelief to be correlated with both ideological and affective polarization.

To investigate this formally, we need to define the average disbelief towards the out-group. We first define the disbelief towards the out-group for each task  $j \in \{1, 2, \dots, 8\}$  as follows:

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

<sup>12</sup>This is consistent with the theoretical prediction of Cheng and Hsiaw (2022) that disagreement in opinions is accompanied by disagreement about the credibility of experts.

where  $p_{i,g(i),j}^g$  is the estimated accuracy rate towards in-group  $g$ , and  $p_{i,g(i),j}^{g'}$  is the 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 disbelief against out-groups.

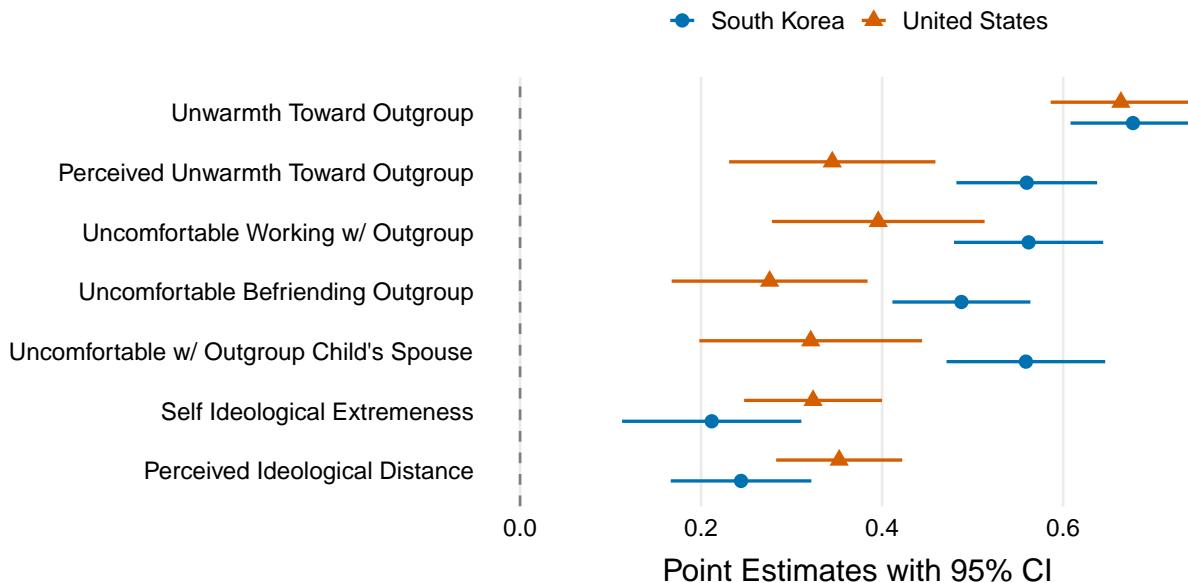
We next define the average of the measure over factual question tasks  $j \in \{1, 2, \dots, 8\}$

$$\text{disbelief}_{i,g(i)} := \frac{1}{8} \sum_{j=1}^8 \text{disbelief}_{i,g(i),j}.$$

We then run bivariate regressions for different polarization measures on  $\text{disbelief}_{i,g(i)}$ .

**Results** Figure 3 presents the correlation between partisan disbelief and various forms of polarization. It shows that all measures of polarization—self-reported affective polarization, meta-perceived affective polarization, self-reported ideological extremeness, and meta-perceived ideological polarization—are positively correlated with partisan disbelief. That is, partisan disbelief is associated with polarization, although the causal direction may run both ways.

Figure 3: Correlation between Partisan Disbelief and Polarization: Baseline Survey



Note: This figure presents the correlations between target-based partisan disbelief and various measures of polarization. Each point shows the estimated coefficient from a bivariate regression of individual-level disbelief on a given polarization measure, with 95% confidence intervals.

### 3 Correcting Partisan Disbelief: Experimental Design

Section 2 documents that partisan disbelief is prevalent in both of the two polarized countries in spite of the difference in institutions and cultures. Given this result, we aim to explore the effect of providing information that different partisan groups are equally knowledgeable in terms of judging whether several non-partisan statements are true or false. Specifically, we explore (1) the effect on disbelief, (2) the effect on in-group bias in information processing, and (3) the effect on affective polarization. We focus on party  $R$  supporters and party  $L$  supporters in South Korea and the U.S..

#### 3.1 Data Collection

We conducted an online survey experiment using Qualtrics in September 2025 in South Korea and the United States. Respondents were recruited through Rakuten Insight in South Korea and PureSpectrum in the United States.<sup>13</sup> We employed quota sampling based on gender and age. Since our focus is on individuals with partisan disbelief, we restricted the sample to respondents who supported either party  $R$  or party  $L$ .<sup>14</sup> Participants received a participation fee from the survey firm. The study was pre-registered with the AEA RCT Registry (AEARCTR-0016557), and the survey protocol was approved by the Institutional Review Board at Sungkyunkwan University (SKKU 2025-08-064-001). To ensure respondent quality, we included a satisficing check at the beginning of the survey. Respondents who failed this check were not allowed to proceed further. In total, 4,262 respondents participated in South Korea and 4,375 in the United States. Based on the power calculation, our target was 4,252 for each country. For both countries, we only use the first 4,252 observations from respondents who completed the survey.

#### 3.2 Survey Structure

The survey flow is given by Figure 4. The main part of this experiment contains nine tasks where respondents are asked to judge whether a statement is true or false:  $j \in \{1, \dots, 9\}$  as in the baseline survey. There are four types of tasks. First, we have two tasks on factual questions before the treatment ( $j = 1, 2$ ). Second, after the treatment, we have three tasks on factual questions ( $j = 3, 4, 5$ ) and two tasks on conspiracy theory questions with partisan signals ( $j = 6, 7$ ). Finally, we have two tasks on factual questions with education group signals ( $j = 8, 9$ ). The main focus of our analysis is tasks on non-partisan facts ( $j = 1, \dots, 5$ ). The list of these tasks is available in Table 4. All of the tasks were used in the baseline survey. Note that the tasks with education group signals serve as a benchmark for gauging the magnitude of the partisan in-group bias in information processing.

---

<sup>13</sup>Since we used PureSpectrum's vendors in the baseline survey, we excluded respondents who completed the baseline survey from entering the experiment

<sup>14</sup>At the beginning of the survey, respondents were asked which party they supported. Those who did not choose either party  $R$  or  $L$  were screened out and could not proceed further. Thus, the sample includes only supporters of party  $R$  or party  $L$ .

Figure 4: Survey Flow of the Experiment

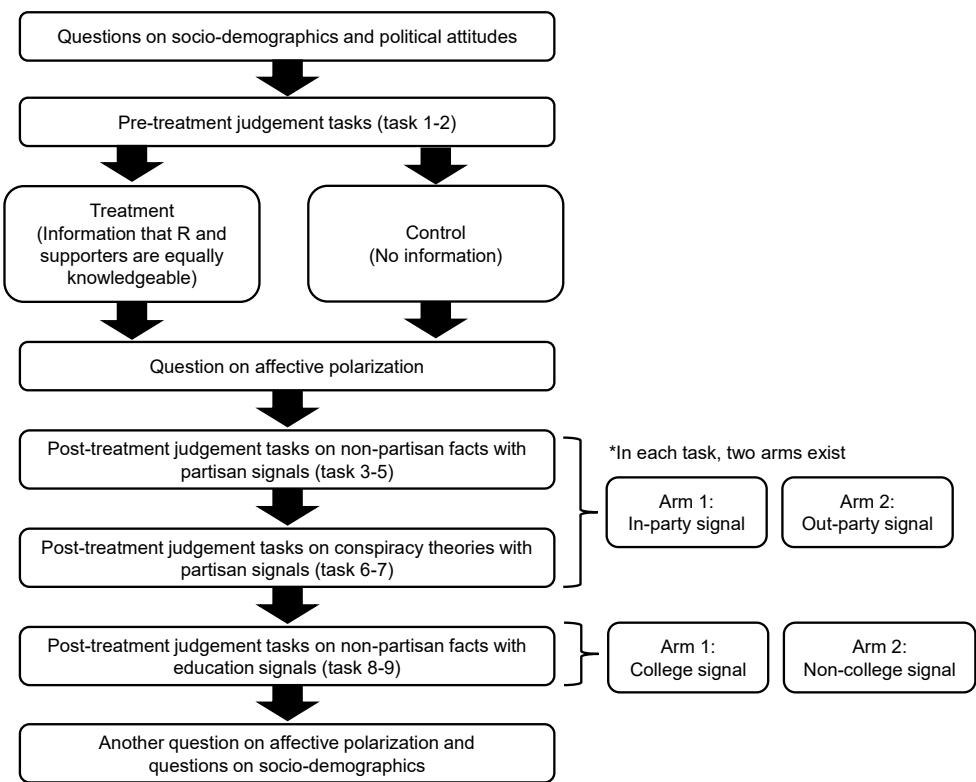


Table 4: List of True-or-False Questions in Experiment

	<b>South Korea</b>	<b>United States</b>
<b>Pre-treatment</b>	The term of the National Assembly is 2 years. (F)	To revise the Constitution, approval from more than three-fourths of the state legislatures is required. (T)
	New Zealand is a country located in the Middle East. (F)	New Zealand is a country located in the Middle East. (F)
<b>Post-treatment</b>	To revise the Constitution, a majority of votes in a national referendum is required. (T)	The country's nominal GDP growth rate in the last year was lower than 7%. (T)
	The iPhone was invented before 2000. (F)	The iPhone was invented before 2000. (F)
	The highest mountain in the country is Hallasan. (T)	The term of office in the Senate is 4 years. (F)

### 3.3 Pre-Treatment Judgment Tasks

After socio-demographic questions, respondents were asked to conduct two true or false judgment tasks on non-partisan facts ( $j = 1, 2$ ). In each task, each respondent was 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. This is exactly the same as in the baseline survey. The objective of these two tasks is to measure each individual's partisan disbelief before the treatment intervention.

### 3.4 Treatment

After this stage, half of the respondents were randomly assigned to the treatment condition and the other half to the control condition. In the control group, respondents were reminded—via an interactive format—of their own estimated accuracy assessments of each party's supporters from the previous two judgment tasks. In the treated group, in addition to this reminder, respondents were also informed that supporters of party  $R$  and party  $L$  were almost equally knowledgeable in those two tasks, based on our baseline survey (specifically, the difference in accuracy rates was less than 5%). Figure 5 presents examples of the information shown in the control and treatment conditions when a respondent thinks that Republican Party supporters are more knowledgeable than Democratic Party supporters.

To ensure that respondents in the treated group carefully read the passage, we included a follow-up true-or-false question. Respondents could not proceed further without selecting the correct answer.<sup>15</sup> If they answered incorrectly, they were prompted to try again until the correct response was chosen.

---

<sup>15</sup>The correct answer in the control group differs depending on respondents.

For respondents who estimated that their in-group's accuracy rate exceeded the out-group's by more than 5%, the treatment information functioned as a corrective, addressing this misperceived partisan disbelief. This design allows us to analyze the effect of correcting partisan disbelief.

### 3.5 Post-Treatment Judgment Tasks

Subsequently, respondents were asked to do three additional judgment tasks about whether a non-partisan fact is true or false ( $j = 3, 4, 5$ ). Each task proceeded as follows.

- (i) Each respondent was 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.
- (ii) Each respondent randomly received 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 (see Figure 6 as an example). We presented the correct answers, which were endorsed by both the majority of  $R$  and the majority of  $L$  supporters in our baseline survey.
- (iii) Respondents were 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. By analyzing how the degree of judgment revisions before and after a signal differs depending on whether the signal is in-group or out-group, we estimated the degree of in-group bias in information processing.

In addition to them, we have another four tasks ( $j = 6, 7$  are tasks on conspiracy theories, and  $j = 8, 9$  are tasks on non-partisan facts with education group signals). These are used in the supplementary analysis (see the Appendix for the details).

### 3.6 Measuring Affective Polarization

In addition, we measured affective polarization using two approaches, as in the baseline survey. The first question asked respondents: "On a scale from 0 to 100, where 0 means very cold or unfavorable feelings and 100 means very warm or favorable feelings, how warmly do you feel toward supporters of each political party?" The second question asked respondents: "How would you feel about being in the following types of relationships with supporters of each political party: (i) colleagues at work, (ii) close friends, and (iii) your own or your child's spouse?" To examine whether the treatment effect on affective polarization quickly diminishes, we administered the first question immediately after the treatment and the second question after the completion of all judgment tasks (see Figure 4).

Figure 5: Treatment and Control (Example)

We asked you to evaluate the average accuracy rates of Republican Party supporters and Democratic Party supporters in the two true-or-false judgment tasks. Across two tasks, you estimated that on average,

- Republican Party supporters correctly judge true-or-false statements with [ ] % accuracy.
- Democratic Party supporters correctly judge true-or-false statements with [ ] % accuracy.

That is, you think that Republican Party supporters are more knowledgeable than Democratic Party supporters.

**Control group:** Please judge whether the following is true or not based on the above. You cannot move to the next question without choosing the correct answer.

You estimated that Republican Party supporters are more knowledgeable than Democratic Party supporters in the two judgment tasks.

- True
- False

**Treated group:** But this is incorrect. In our earlier survey, we found that Republican Party supporters and Democratic Party supporters demonstrated nearly the same level of knowledge on the issue. To be specific, for each of the two true-or-false questions we asked, the difference in the percentage of correct answers between the two groups was consistently less than 5%.

Please judge whether the following is true or not based on the above. You cannot move to the next question without choosing the correct answer.

Republican Party supporters and Democratic Party supporters are nearly equally knowledgeable across the two judgment tasks (the difference in correct response rates between the two groups was consistently less than 5%).

- True
- False

Figure 6: Signal (Example)

Your judgment in the previous page is that "The country's nominal GDP growth rate last year was lower than 7%" is TRUE.

According to our previous survey, **a majority of Republican Party supporters** also say that "The country's nominal GDP growth rate last year was lower than 7%" is **TRUE**.

**A majority of Republican Party supporters**  
say that  
"The country's nominal GDP growth rate last  
year was lower than 7%" is **TRUE**.



Please choose the appropriate sentence based on the above. You cannot move to the next question without choosing the correct answer.

- My initial judgment was different from the judgment by a majority of Republican Party supporters.
- My initial judgment was the same as the judgment by a majority of Republican Party supporters.

Table 5: Correlation between Attrition and Treatment: Experiment

	(1) SK	(2) US
Treatment	-0.002 (0.009)	0.004 (0.008)
Observations	4734	4778
Mean of outcome	0.099	0.084

Note: This table shows the correlation between treatment assignment and survey attrition among respondents who passed the attention check at the beginning of the survey. Each coefficient is estimated from a regression of an attrition dummy on the treatment indicator. Standard errors are reported in parentheses.

### 3.7 Sample Selection

In total, 4,734 respondents in South Korea and 4,778 respondents in the United States passed the attention check at the beginning of the survey. Among them, 4,262 respondents in South Korea and 4,375 in the United States completed the entire survey. The corresponding attrition rates are 0.099 and 0.084, respectively.

Table 5 reports the relationship between attrition and treatment, estimated by regressing an attrition dummy on the treatment indicator among respondents who passed the attention check. The results indicate that the treatment did not increase attrition in either country.

Then, we restrict samples with the following criteria about the prior level of disbelief. To formally define the criteria, for each individual  $i$  in group  $g \in \{R, L\}$  and task  $j$ , we define the (target-based) disbelief on fact  $j$  as

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

where  $p_{i,g(i),j}^g$  is the estimated accuracy rate towards in-group  $g$ , and  $p_{i,g(i),j}^{g'}$  is the 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 disbelief against out-groups.

Then, the pre-treatment disbelief is given by

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

In the following analysis, we restrict our attention to those with  $\text{disbelief}_{i,g(i)}^{pre} > 0.05$ . The treatment provides information that the difference in the accuracy rate is less than 5%. Thus, the treatment is expected to reduce disbeliefs only among those with high enough  $\text{disbelief}_{i,g(i)}^{pre}$ . This sample selection resulted in 2,305 respondents in South Korea and 2,792 in the United States.

## 4 Correcting Partisan Disbelief: Hypotheses

First, we hypothesize that the treatment decreases disbelief in the out-group's performance in the post-treatment judgment tasks.

**Hypothesis 1. (Treatment effect on disbelief).** The post-treatment disbelief regarding non-partisan facts in tasks  $j = 3, 4, 5$  is smaller in the treated group than in the control group.

Importantly, the treatment conveyed only that the accuracy rates in the pre-treatment judgment tasks were nearly identical across partisan groups. It remains logically possible that, although the accuracy rates were identical in the pre-treatment tasks, in-group members could outperform out-group members in the post-treatment tasks. Thus, whether the treatment reduces disbelief in the post-treatment tasks is not evident *ex ante*. This hypothesis should therefore not be interpreted as a mere manipulation check of the treatment. Rather, it tests whether correcting disbelief in knowledge about specific issues can generate a spillover effect on disbelief in knowledge about other issues. Because it is impossible to address all possible issues through correction, examining this hypothesis is important.

The next hypothesis concerns in-group bias in information processing. Suppose an individual encounters information provided by a supporter of the opposing party. If partisan disbelief is present, the individual deems this information unreliable solely because it comes from someone perceived to have low knowledge due to their party affiliation. Therefore, in-group bias in information processing may exist even for non-partisan issues. Therefore, its existence is expected in the control group. Furthermore, the treatment is expected to reduce this bias because it would mitigate partisan disbelief from Hypothesis 1. These considerations lead to the following two hypotheses:

**Hypothesis 2. (In-group bias in information processing).** Partisans have an in-group bias in information processing for non-partisan facts in the control group.

**Hypothesis 3. (Treatment effect on in-group bias in information processing).** Partisans have a smaller in-group bias in information processing for non-partisan facts in the treated group than in the control group.

Lastly, when individuals perceive out-group members as having low levels of knowledge, they would develop greater distrust toward them. Therefore, partisan disbelief may exacerbate affective polarization.<sup>16</sup> Therefore, we expect that the treatment reduces affective polarization:

**Hypothesis 4. (Treatment effect on affective polarization).** The treatment decreases the affective polarization.

In the following, we test these three hypotheses based on our experiment.

---

<sup>16</sup>While he does not consider partisan disbelief, Stone (2020) theoretically demonstrate that a combination of three types of misperceptions (a prior bias against the other agent's character, the false consensus bias, and limited strategic thinking) creates affective polarization. Furthermore, Bowen et al. (2023) theoretically demonstrate that misperceptions about friends' sharing on social media induce opinion polarization.

## 5 Correcting Partisan Disbelief: Analysis

This section presents the empirical specifications used to evaluate the hypotheses and the corresponding results.

### 5.1 Summary Statistics and Balance Check: Experiment

Table 6 presents the summary statistics of respondents with disbelief $_{i,g(i)}^{pre} > 0.05$ . Although the female ratio is slightly different between the treated and control groups in the U.S., there are no statistically significant differences between the groups in other variables. Overall, the randomization worked well.

Table 6: Summary Statistics: Experiment

	SK Treated	SK Control	SK Difference	US Treated	US Control	US Difference
Right-wing Party Supporters Ratio	0.256	0.250	0.006 (0.018)	0.543	0.519	0.024 (0.019)
Female Ratio	0.481	0.482	-0.001 (0.021)	0.488	0.532	-0.044 (0.019)
College-educated Ratio	0.804	0.795	0.010 (0.017)	0.541	0.528	0.013 (0.019)
Age (50+) Ratio	0.709	0.711	-0.002 (0.019)	0.669	0.695	-0.026 (0.018)
Pre-treatment Accuracy Rate	0.955	0.958	-0.003 (0.006)	0.848	0.834	0.014 (0.009)
Pre-treatment Out-group Disbelief	0.350	0.343	0.008 (0.011)	0.341	0.336	0.005 (0.010)
Observations	1166	1139		1385	1407	

Note: This table reports summary statistics for the experimental sample in South Korea (SK) and the United States (U.S.). Columns show means for treated and control groups, along with their differences and standard errors in parentheses.

### 5.2 Effect on Partisan Disbelief

We start with testing Hypothesis 1: the treatment effect on partisan disbelief.

**Measurement** To test the hypothesis, we need to define the measurement of disbelief about non-partisan facts after receiving the treatment. As in the case of the ex-ante disbelief, this ex-post disbelief is measured by

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

We use this measurement to test the hypothesis.

**Specification for Hypothesis 1** To test Hypothesis 1, we run the following regression:

$$\text{disbelief}_{i,g(i)}^{post,f} = \alpha T_i + \text{const.} + \varepsilon_i, \quad (4)$$

where  $T_i = 1$  if individual  $i$  is treated. Given this, we expect  $\hat{\alpha} < 0$  as the empirical specification of Hypothesis 1.

**Results** Table 7 presents the results. The treatment reduced disbelief by 5 percentage points in South Korea and 7.6 percentage points in the United States, both statistically significant effects. Nevertheless, partisan disbelief did not disappear entirely. The mean level of partisan disbelief is 0.232 (23.2%) in South Korea and 0.195 (19.5%) in the United States. Thus, the treatment reduced only part of the overall disbelief. Overall, Hypothesis 1 was supported.

Table 7: Treatment Effects on Partisan Disbelief

	(1) SK	(2) US
Treatment	-0.050 (0.010)	-0.076 (0.009)
Observations	2305	2792
Mean of outcome	0.232	0.195

Note: This table reports the estimated treatment effects on post-treatment partisan disbelief in South Korea (SK) and the United States (U.S.). The dependent variable is the average difference in perceived accuracy between in-group and out-group supporters across five non-partisan factual statements. Each coefficient represents the effect of being assigned to the treatment group ( $T_i = 1$ ) relative to the control group. Standard errors, shown in parentheses, are robust to heteroskedasticity.

### 5.3 In-Group Bias in Information Processing

Next, we test Hypotheses 2 and 3: the presence of in-group bias in the control group and the treatment effect on this bias. The key idea for measuring in-group bias is as follows. For all non-partisan statements, the majority's judgment was identical; hence, the informational content of the signals was constant, with only the source identity varying. Accordingly, if respondents revised their answers toward the correct response more frequently when the signal came from the in-group than when it came from the out-group, we interpret this as evidence of in-group bias in information processing.

**Measurement** To formalize this idea, 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

$$\mu_{i,j,0} = \begin{cases} \frac{a_{i,j,0}}{100} & \text{if } J_{i,j,0} = 1 \\ 1 - \frac{a_{i,j,0}}{100} & \text{if } J_{i,j,0} = 0 \end{cases} \quad (5)$$

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

$$\mu_{i,j,1} = \begin{cases} \frac{a_{i,j,1}}{100} & \text{if } J_{i,j,1} = 1 \\ 1 - \frac{a_{i,j,1}}{100} & \text{if } J_{i,j,1} = 0 \end{cases} \quad (6)$$

$(J_{ij0}, J_{ij1})$  and  $(\mu_{ij0}, \mu_{ij1})$  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",  $y_{i,j}^J \in \{0, 1\}$ , where  $y_{i,j}^J = 1$  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",  $y_{i,j}^\mu \in \{0, 1\}$ , where  $y_{i,j}^\mu = 1$  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** 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  $y_{i,j} = \{y_{i,j}^J, y_{i,j}^\mu\}$  be the *change* in respondent  $i$ 's opinion on the same task  $j$  before and after the signals. We estimate the following separately for the treated group ( $T_i = 1$ ) and the control group ( $T_i = 0$ ) for the identical task  $j$ .

$$y_{i,j} = \beta \mathbb{1}\{s_{i,j} = I\} + \eta_j + \varepsilon_{i,j} \quad (7)$$

We denote the estimands of  $\beta$  for the treated group  $\beta^T$  and the control group  $\beta^C$ .

We use both measures  $(y_{i,j}^J, y_{i,j}^\mu)$  as  $y_{i,j}$ . In the post-teratment judgement tasks on non-partisan 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,  $\beta$  represents the degree of in-group bias in information processing. Therefore, we expect that  $\hat{\beta} > 0$  in (7) holds in the control group as Hypothesis 2.

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

**Results for Hypothesis 2** Table 8 reports the results on in-group bias in the control group. In South Korea, in-group signals prompted respondents to revise their opinions toward the correct choice more frequently than out-group signals. Specifically, in-group signals increased the probability of a “dummy update” by 5.8 percentage points and a “continuous update” by 7.8 percentage points, both statistically significant effects. That is, in-group bias in information processing exists even for non-partisan issues.

It is worth noting that, on average, only 10% of respondents engaged in dummy updates. In our judgment tasks, a majority of respondents initially selected the correct answer; for them, the signals merely confirmed their prior judgment, and thus no dummy update could occur. Hence, the relatively low rate of dummy updates is unsurprising. Moreover, the effect size was larger for continuous updates than for binary updates, which is consistent with the same logic: binary updates are inherently less likely to occur than continuous updates.

Table 8: In-Group Bias in Information Processing (Control Group)

	(1) SK Dummy	(2) SK Continuous	(3) US Dummy	(4) US Continuous
In-Group Signal	0.058 (0.010)	0.078 (0.008)	0.007 (0.009)	0.041 (0.012)
Observations	3417	3417	4221	4221
Mean of outcome	0.102	0.424	0.170	0.481

Note: this table reports the estimated in-group bias in information processing for the control group in South Korea (SK) and the United States (U.S.). The dependent variable is the change in respondents’ judgments before and after receiving a signal, measured as either a binary (“dummy”) update or a continuous update. The regressor is an indicator of whether the signal originates from in-group members. The specifications control for task fixed effects. Standard errors, shown in parentheses, are robust to heteroskedasticity.

The presence of in-group bias in the United States is more nuanced. In-group signals increased the probability of a continuous update by 4.1 percentage points. Although this effect is smaller than that observed in South Korea, it is statistically significant. By contrast, in-group signals increased the probability of a dummy update by only 0.7 percentage points, which was not statistically significant. Thus, evidence of in-group bias is found only for continuous updates. These results suggest that in-group bias in information processing is stronger in South Korea than in the United States. This is consistent with the fact that the average of partisan disbelief is larger in South Korea than in the U.S. (Table 7).

Overall, Hypothesis 2 is fully supported in South Korea and partly supported in the U.S..

**Specification for Hypothesis 3** Having the result in hand, we next test Hypothesis 3: the treatment effect on in-group bias. For this, we interact the in-group signal dummy with the treatment

as follows.

$$y_{i,j} = \beta_1 \mathbb{1}\{s_{i,j} = I\} + \beta_2 T_i + \beta_3 (\mathbb{1}\{s_{i,j} = I\} \times T_i) + \eta_j + \varepsilon_{i,j} \quad (8)$$

Hypothesis 3 predicts  $\hat{\beta}_3 < 0$ .

**Results for Hypothesis 3** Table 9 shows the results. In South Korea, estimates of  $\hat{\beta}_3$  were negative and statistically significant for both binary and continuous measures, indicating that the treatment reduced in-group bias. The effect size was substantial: in both cases, the treatment reduced more than half of the in-group bias. For example, in the control group, in-group signals increased the probability of a continuous update by 7.8 percentage points, whereas in the treated group, the increase was only 1.8 percentage points. This indicates that partisan disbelief has a causal effect on in-group bias in information processing, and that correcting disbelief through information provision is highly effective in reducing it.

Table 9: Treatment Effects on In-Group Bias in Information Processing

	(1) SK Dummy	(2) SK Continuous	(3) US Dummy	(4) US Continuous
In-Group Signal	0.058 (0.010)	0.078 (0.009)	0.007 (0.010)	0.041 (0.012)
Treatment	0.016 (0.009)	0.023 (0.019)	-0.018 (0.008)	0.023 (0.019)
In-Group Signal x Treatment	-0.031 (0.002)	-0.060 (0.028)	0.018 (0.007)	-0.036 (0.015)
Observations	6915	6915	8376	8376
Mean of outcome	0.103	0.421	0.165	0.484

Note: This table reports the treatment effects on in-group bias in information processing for South Korea (SK) and the United States (U.S.). The dependent variable measures the change in respondents' judgments before and after receiving a signal, defined either as a binary ("dummy") update or a continuous update. The key variable *In-Group Signal x Treatment* captures the differential effect of in-group versus out-group signals for treated respondents relative to the control group. Task fixed effects are included, and robust standard errors are shown in parentheses.

In the U.S., the results are more nuanced because in-group bias was not observed for binary updates even in the control group. First, in the case of continuous updates, the result is consistent with what we expected. The estimate of  $\hat{\beta}_3$  was negative and statistically significant. In the control group, in-group signals increased the probability of a continuous update by 4.1 percentage points, whereas in the treated group, the increase was only 0.5 percentage points. Therefore, the treatment was highly effective. However, for discrete updates, we did not obtain such a result because in-group bias was not observed even in the control group. Although the effect size was much smaller, the estimate of  $\hat{\beta}_3$  was not negative but positive. This may suggest a backfire effect of the treatment when the initial level of in-group bias is negligible.<sup>17</sup>

<sup>17</sup>Several papers show a possibility that belief correction or priming has a backfire effect (e.g., Nyhan and Reifler,

Overall, Hypothesis 3 is fully supported in South Korea and partly supported in the U.S..

## 5.4 Affective Polarization

Finally, we test Hypothesis 4: the treatment effect on affective polarization.

**Measurement** For this purpose, we need to measure respondents' unfavorable feelings toward out-party members and in-party members. The difference between them is:

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

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

Favorable feelings were measured in two ways. First, just after the treatment information, we asked respondents to rate positive feelings toward out-party members and in-party members from 0 to 100. We normalize this measure to range from 0 to 1 and denote  $\text{pol}_{i,g(i)}$  using this measurement by  $\text{unfav}_{i,g(i)}$ . Second, after the completion of all judgment tasks, we asked 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. We denote this by  $\text{uncomf}_{i,g(i)}$ .

**Specification for Hypothesis 4** To test the hypothesis, we estimate the following:

$$\text{pol}_{i,g(i)} = \text{const.} + \gamma T_i + \varepsilon_i. \quad (9)$$

Then, Hypothesis 4 predicts  $\hat{\gamma} < 0$ .

**Results** Table 10 shows the results. First, columns (1) and (3) present the short-run effects on affective polarization, as this measure was collected immediately after the treatment. The treatment reduced affective polarization by 2.5 percentage points in South Korea and by 8.1 percentage points in the United States, both statistically significant effects.

Columns (2) and (4) present the effects on this second measure, showing that the treatment effects were not statistically significant, although the point estimates remained negative.

Note that the contents of the questions, as well as their timing, differ across the two measures. Given that the two measures are highly correlated, the difference in timing is likely to matter more than the difference in content.

---

2010; Bicchieri and Dimant, 2022; Colonnelli et al., 2024).

Table 10: Treatment Effects on Affective Polarization

	(1) SK Unfav	(2) SK Uncomf	(3) US Unfav	(4) US Uncomf
Treatment	-0.025 (0.012)	-0.002 (0.012)	-0.081 (0.013)	-0.023 (0.016)
Observations	2305	2305	2792	2792
Mean of outcome	0.528	0.433	0.493	0.271

Note: This table reports the estimated treatment effects on affective polarization in South Korea (SK) and the United States (U.S.). The dependent variable is the difference in favorable feelings toward in-party versus out-party members, measured either by self-reported warmth ratings (*Unfav*) or by (un)comfortableness in social relationships with out-party members (*Uncomf*). Each coefficient represents the effect of being assigned to the treatment group ( $T_i = 1$ ) relative to the control group. Robust standard errors are reported in parentheses.

## 5.5 Additional Analysis

The Online Appendix provides supplementary analyses of the experiment. First, although previous research shows that experimenter demand effects are limited in online survey experiments (Mummolo and Peterson, 2019), it remains important to verify that our findings are not driven by such effects. To this end, we reanalyzed the data after excluding respondents who appeared to pander to the hypothesis implied by the experimenter. The results remain unchanged, implying that experimenter demand effects are not a serious concern in our setting.

Second, the experiment also included two additional judgment tasks involving conspiracy theories. Contrary to our expectation, we did not find evidence of in-group bias in information processing for these tasks. One possible explanation is that, for highly partisan issues such as conspiracy beliefs, a signal indicating that a majority of a party’s supporters endorse or reject a conspiracy theory may not be interpreted literally by respondents.

Third, the experiment included two additional judgment tasks on non-partisan issues in which each respondent randomly received one of two signals: one indicating the majority opinion among college graduates, and the other indicating the majority opinion among non-college graduates. Respondents might update their opinions differently depending on whether the signal originated from college graduates. This serves as a benchmark for gauging the magnitude of the partisan in-group bias in information processing observed in the main analysis. We found that the partisan in-group bias is substantially larger than this education-based bias, suggesting that partisan in-group bias cannot simply be attributed to misperceptions about the educational composition of each political group. If partisan disbelief merely reflected such misperceptions, the effect of partisan signals should not exceed that of education-based signals.

## 6 Conclusion

This paper introduced and empirically validated a new concept—*partisan disbelief in knowledge*: the belief that one’s in-group is more knowledgeable than the opposing party, even about basic, non-partisan facts. Across large baseline surveys in South Korea and the United States, we show that this disbelief is widespread: partisans perceive higher accuracy among co-partisans than among the out-party by roughly 15 percentage points or more when judging true–false statements about non-partisan facts.

Two survey experiments clarify why this matters. Partisan disbelief distorts information processing even outside explicitly political domains. Crucially, simple corrective evidence that both sides are similarly knowledgeable reduces partisan disbelief and dampens the resulting in-group bias. These corrections also lower affective polarization, albeit transiently.

Taken together, the findings shift the focus of polarization from differences in attitudes alone to differences in perceived competence. We identify a cognitive mechanism—rooted in social identity—that obstructs information exchange and mutual understanding. The results also point to actionable remedies: interventions that normalize perceptions of cross-party competence can improve how people interpret factual information and, in the short run, soften affective divides. Future work should test how to sustain these gains over time and at scale, and examine whether complementary interventions (e.g., repeated exposure, messenger choice, or institutional cues) can entrench more accurate beliefs about out-party competence while preserving open disagreement on values.

## References

- Ahler, D. J. (2014). Self-fulfilling misperceptions of public polarization. *The Journal of Politics* 76(3), 607–620.
- Ahler, D. J. and G. Sood (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics* 80(3), 964–981.
- Alesina, A., A. Miano, and S. Stantcheva (2020). The polarization of reality. *AEA Papers and Proceedings* 110, 324–328.
- Alesina, A., S. Stantcheva, and E. Teso (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review* 108(2), 521–554.
- Angrisani, M., A. Guarino, P. Jehiel, and T. Kitagawa (2021). Information redundancy neglect versus overconfidence: A social learning experiment. *American Economic Journal: Microeconomics* 13(3), 163–197.
- Bicchieri, C. and E. Dimant (2022). Nudging with care: The risks and benefits of social information. *Public choice* 191(3), 443–464.

- Bowen, T. R., D. Dmitriev, and S. Galperti (2023). Learning from shared news: When abundant information leads to belief polarization. *The Quarterly Journal of Economics* 138(2), 955–1000.
- Boxell, L., M. Gentzkow, and J. M. Shapiro (2024). Cross-country trends in affective polarization. *Review of Economics and Statistics* 106(2), 557–565.
- Bursztyn, L. and D. Y. Yang (2022). Misperceptions about others. *Annual Review of Economics* 14(1), 425–452.
- Cappelen, A. W., B. Enke, and B. Tungodden (2025). Universalism: global evidence. *American Economic Review* 115(1), 43–76.
- Cheng, H. and A. Hsiaw (2022). Distrust in experts and the origins of disagreement. *Journal of Economic Theory* 200, 105401.
- Cheong, Y. and S. Haggard (2023). Political polarization in korea. *Democratization* 30(7), 1215–1239.
- Chopra, F., I. Haaland, and C. Roth (2024). The demand for news: Accuracy concerns versus belief confirmation motives. *The Economic Journal* 134(661), 1806–1834.
- Colonnelly, E., N. J. Gormsen, and T. McQuade (2024). Selfish corporations. *Review of Economic Studies* 91(3), 1498–1536.
- Cox, L., P. Cubillos, and C. Le Foulon (2025). Affective polarization and democratic erosion: evidence from a context of weak partisanship. *Political Science Research and Methods*, 1–8.
- De Filippis, R., A. Guarino, P. Jehiel, and T. Kitagawa (2022). Non-bayesian updating in a social learning experiment. *Journal of Economic Theory* 199, 105188.
- Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American Economic Review* 112(3), 899–927.
- Dias, N. C., Y. Lelkes, and J. Pearl (2025). American partisans vastly under-estimate the diversity of other partisans' policy attitudes. *Political Science Research and Methods* 13(3), 725–735.
- Dimant, E. (2024). Hate trumps love: The impact of political polarization on social preferences. *Management Science* 70(1), 1–31.
- Dimant, E., F. Galeotti, and M. C. Villeval (2024). Motivated information acquisition and social norm formation. *European Economic Review* 167, 104778.
- Druckman, J. N., S. Klar, Y. Krupnikov, M. Levendusky, and J. B. Ryan (2022). (mis) estimating affective polarization. *The Journal of Politics* 84(2), 1106–1117.
- Druckman, J. N. and M. S. Levendusky (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly* 83(1), 114–122.

- Enke, B., R. Rodríguez-Padilla, and F. Zimmermann (2023). Moral universalism and the structure of ideology. *The Review of Economic Studies* 90(4), 1934–1962.
- Faia, E., A. Fuster, V. Pezone, and B. Zafar (2024). Biases in information selection and processing: Survey evidence from the pandemic. *Review of Economics and Statistics* 106(3), 829–847.
- Fang, X., S. Heuser, and L. S. Stötzer (2025). How in-person conversations shape political polarization: Quasi-experimental evidence from a nationwide initiative. *Journal of Public Economics* 242, 105309.
- Fowler, J. H. and C. D. Kam (2007). Beyond the self: Social identity, altruism, and political participation. *The Journal of Politics* 69(3), 813–827.
- Fryer Jr, R. G., P. Harms, and M. O. Jackson (2019). Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. *Journal of the European Economic Association* 17(5), 1470–1501.
- Gidron, N., J. Adams, and W. Horne (2020). *American Affective Polarization in Comparative Perspective*. Elements in American Politics. Cambridge University Press.
- Gidron, N., J. Adams, and W. Horne (2023). Who dislikes whom? affective polarization between pairs of parties in western democracies. *British Journal of Political Science* 53(3), 997–1015.
- Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics* 127(3), 1287–1338.
- Green, D. P., B. Palmquist, and E. Schickler (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.
- Hill, S. J. (2017). Learning together slowly: Bayesian learning about political facts. *The Journal of Politics* 79(4), 1403–1418.
- Huber, G. A. and N. Malhotra (2017). Political homophily in social relationships: Evidence from online dating behavior. *The Journal of Politics* 79(1), 269–283.
- Hudde, A., P. Jungeilges, C. Proch, and T. Schmitt (2024). How warm are political interactions? a new measure of affective fractionalization. *PLOS ONE* 19(5), e0294401.
- Huddy, L., L. Mason, and L. Aarøe (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review* 109(1), 1–17.
- Huddy, L. and O. Yair (2021). Reducing affective polarization: Warm group relations or policy compromise? *Political Psychology* 42(2), 291–309.
- Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science* 22(1), 129–146.

- Iyengar, S. and S. J. Westwood (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* 59(3), 690–707.
- Kashner, D. and M. Stalinski (2024). Preempting polarization: An experiment on opinion formation. *Journal of Public Economics* 234, 105122.
- Kishishita, D. and T. Matsumoto (2024). Self-benefits, fiscal risk, and political support for the public healthcare system. *European Journal of Political Economy* 85, 102597.
- Lees, J. and M. Cikara (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour* 4(3), 279–286.
- Levendusky, M. S. (2018). Americans, not partisans: Can priming american national identity reduce affective polarization? *The Journal of Politics* 80(1), 59–70.
- Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review* 111(3), 831–870.
- Little, A. T. (2019). The distortion of related beliefs. *American Journal of Political Science* 63(3), 675–689.
- Lord, C. G., L. Ross, and M. R. Lepper (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37(11), 2098.
- Mill, W. and J. Morgan (2022). The cost of a divided america: an experimental study into destructive behavior. *Experimental Economics* 25(3), 974–1001.
- Moorthy, A. (2025). Whom do we learn from? Available at SSRN: <https://ssrn.com/abstract=5190331>.
- Mummolo, J. and E. Peterson (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review* 113(2), 517–529.
- Nyhan, B. and J. Reifler (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior* 32(2), 303–330.
- Peterson, E., S. Goel, and S. Iyengar (2021). Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political Science Research and Methods* 9(2), 242–258.
- Shafranek, R. M. (2021). Political considerations in nonpolitical decisions: A conjoint analysis of roommate choice. *Political Behavior* 43(1), 271–300.
- Silver, L. et al. (2022, November 16). Most across 19 countries see strong partisan conflicts in their society, especially in south korea and the u.s. Pew Research Center, Short Reads. Accessed: 2025-10-09.

- Stone, D. F. (2020). Just a big misunderstanding? bias and bayesian affective polarization. *International Economic Review* 61(1), 189–217.
- Taber, C. S., D. Cann, and S. Kucsova (2009). The motivated processing of political arguments. *Political Behavior* 31(2), 137–155.
- Taber, C. S. and M. Lodge (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50(3), 755–769.
- Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics* 16(2), 1–38.
- Voelkel, J. G., J. Chu, M. N. Stagnaro, J. S. Mernyk, C. Redekopp, S. L. Pink, J. N. Druckman, D. G. Rand, and R. Willer (2023). Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes. *Nature Human Behaviour* 7(1), 55–64.
- Whitt, S., A. B. Yanus, B. McDonald, J. Graeber, M. Setzler, G. Ballingrud, and M. Kifer (2021). Tribalism in america: Behavioral experiments on affective polarization in the trump era. *Journal of Experimental Political Science* 8(3), 247–259.
- Zhang, Y. and D. G. Rand (2023). Sincere or motivated? partisan bias in advice-taking. *Judgment and Decision Making* 18, e29.
- Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review* 110(2), 337–363.

# **Online Appendices for Distrusting the Out-Party: Partisan Disbelief and Biased Information Processing.” (Not for Publication)**

## **Contents**

---

<b>A Perceiver-based Partisan Disbelief</b>	<b>A2</b>
<b>B Baseline Survey: Partisan Disbelief for Conspiracy Theory Questions</b>	<b>A2</b>
<b>C Baseline Survey: Partisan Disbelief for Each Issue</b>	<b>A5</b>
<b>D Experiment: Conspiracy Theory Questions</b>	<b>A9</b>
D.1 Overview . . . . .	A9
D.2 Treatment Effects on Disbelief . . . . .	A9
D.3 Information Processing . . . . .	A10
<b>E Experiment: Education Signals</b>	<b>A13</b>
<b>F Experiment: Experimenter Demand Effect</b>	<b>A16</b>
<b>G Questionnaires</b>	<b>A19</b>
G.1 Baseline Survey in South Korea . . . . .	A19
G.2 Baseline Survey in the U.S. . . . .	A19
G.3 Experiment in South Korea . . . . .	A19
G.4 Experiment in the U.S. . . . .	A19

---

## A Perceiver-based Partisan Disbelief

In the main text, we use target-based partisan disbelief. Here, we instead define perceiver-based partisan disbelief. Citizens have perceiver-based partisan disbelief if the following two conditions are satisfied:

- (a) Supporters of each political party believe that members of their own party are more knowledgeable than the opposing party's supporters believe them to be. Formally, for each target group  $t \in \{R, L\}$  separately, we estimate

$$p_{i,j}^t = \beta_3 \mathbb{1}\{i = t\} + \mu_j + \varepsilon_{i,j}^t, \quad (10)$$

for group  $g(i) \in \{R, L\}$ . A positive  $\beta_3$  indicates that partisans rate their own group as more knowledgeable than out-partisans do, and (a) expects  $\beta_3 > 0$ .

- (b) By contrast, supporters of both parties and non-partisans are expected to hold similar views about how knowledgeable non-partisans are. Formally, for the target group  $t = N$ , we estimate

$$p_{i,j}^t = \beta_4 \mathbb{1}\{g(i) = R\} + \beta_5 \mathbb{1}\{g(i) = L\} + \mu_j + \varepsilon_{i,j}^t \quad (11)$$

If perceptions of non-partisans do not differ by perceiver type, (b) expects  $\beta_4 = \beta_5 = 0$ .

Perceiver-based partisan disbelief captures cross-perceiver disagreement in evaluations of partisan knowledge.

**Results** Table A1 reports the results of estimating equations (10) and (11). Each column is defined by a country and a target group (RP, LP, or NP), while the rows compare how different perceiver groups rate that same target. For RP targets, RP perceivers assign higher perceived accuracy than LP perceivers by 19.6 percentage points in South Korea and 19.7 percentage points in the United States. For LP targets, LP perceivers assign higher perceived accuracy than RP perceivers by 15.4 percentage points in South Korea and 12.9 percentage points in the United States. For NP targets, perceived accuracy varies less across perceiver groups; in the U.S., however, the corresponding coefficients are positive and statistically significant (0.052 for RP perceivers and 0.045 for LP perceivers). Overall, the table indicates systematic differences in perceived knowledge across perceivers for partisan targets in both countries, consistent with perceiver-based partisan disbelief.

## B Baseline Survey: Partisan Disbelief for Conspiracy Theory Questions

In the baseline survey, we asked respondents about conspiracy theories as well as non-partisan facts (see Table A2 for the list of true-or-false questions about conspiracy theories). This section reports partisan disbelief for conspiracy theory questions.

Table A1: Perceiver-based Partisan Disbelief: Baseline Survey

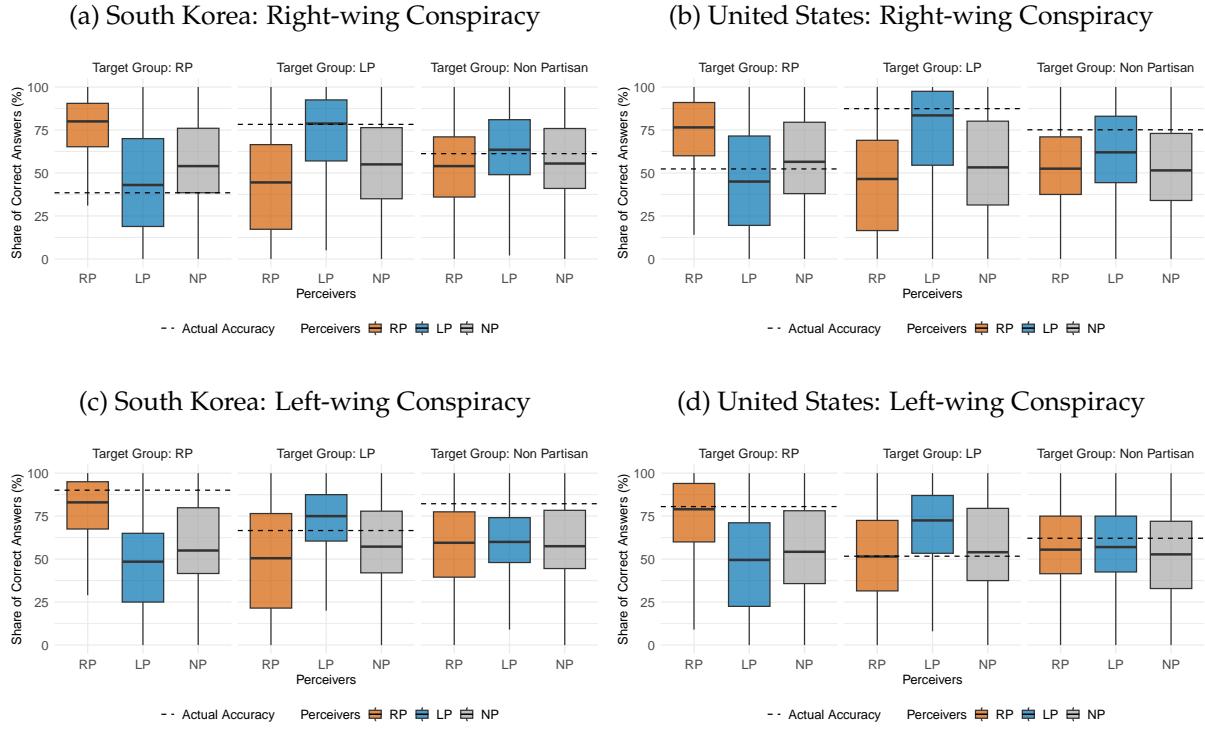
Country	SK RP (1)	SK LP (2)	SK NP (3)	US RP (4)	US LP (5)	US NP (6)
Perceiver = RP	0.196 (0.012)		0.005 (0.018)	0.197 (0.012)		0.052 (0.015)
Perceiver = LP		0.154 (0.015)	0.020 (0.015)		0.129 (0.011)	0.045 (0.015)
Mean of Outcome	0.633	0.663	0.604	0.603	0.604	0.547
Observations	7512	7512	11176	9704	9704	12776

Note: This table reports the results of estimating equations (10) and (11) that test for perceiver-based partisan disbelief in South Korea (SK) and the United States (U.S.). Columns (1)–(3) use the SK sample and columns (4)–(6) use the U.S. sample. Each column corresponds to a target group  $t \in \{\text{RP}, \text{LP}, \text{NP}\}$ , where RP denotes supporters of the right party, LP denotes supporters of the left party, and NP denotes non-partisans. The dependent variable  $p_{i,j}^t$  is perceiver  $i$ 's assessment of the accuracy (probability correct) of target group  $t$  on issue  $j$ . The rows "Perceiver = RP" and "Perceiver = LP" indicate the perceiver group whose ratings are being compared to the omitted perceiver group in that column. All regressions include individual fixed effects and issue fixed effects. Standard errors clustered at the individual level are shown in parentheses.

Table A2: List of True-or-False Questions on Conspiracy Theories

	South Korea	United States
Right-wing	<p>There was widespread election fraud in the parliamentary elections of 2020 and 2024.</p> <p>China is systematically infiltrating major institutions in South Korea to undermine democracy and sovereignty.</p>	<p>The Democratic Party stole the 2020 presidential election.</p> <p>Climate change is a hoax created to push socialist policies and destroy American industry.</p>
Left-wing	<p>The Supreme Court colluded with Yoon Suk-yeol and decided to disqualify Lee Jae-myung from the presidential election.</p> <p>The U.S. government controls major political decisions in Korea, such as suppressing opposition parties under conservative governments.</p>	<p>The Republican administration initiated the Iraq War for oil.</p> <p>The Republicans stole the 2024 presidential election.</p>

Figure A7: Partisan Disbelief by Conspiracy Theory Questions



*Note:* These figures display the distributions of perceived accuracy for conspiracy theory questions by perceiver and target groups in South Korea and the United States. Panels (a) and (b) correspond to right-wing conspiracy items, while panels (c) and (d) correspond to left-wing conspiracy items. The boxes show the interquartile range (25th–75th percentiles), with medians indicated by black lines. Dashed lines represent the actual accuracy rates for each question set. Colors denote perceiver groups: blue for right-party supporters (RP), orange for left-party supporters (LP), and gray for non-partisans (NP).

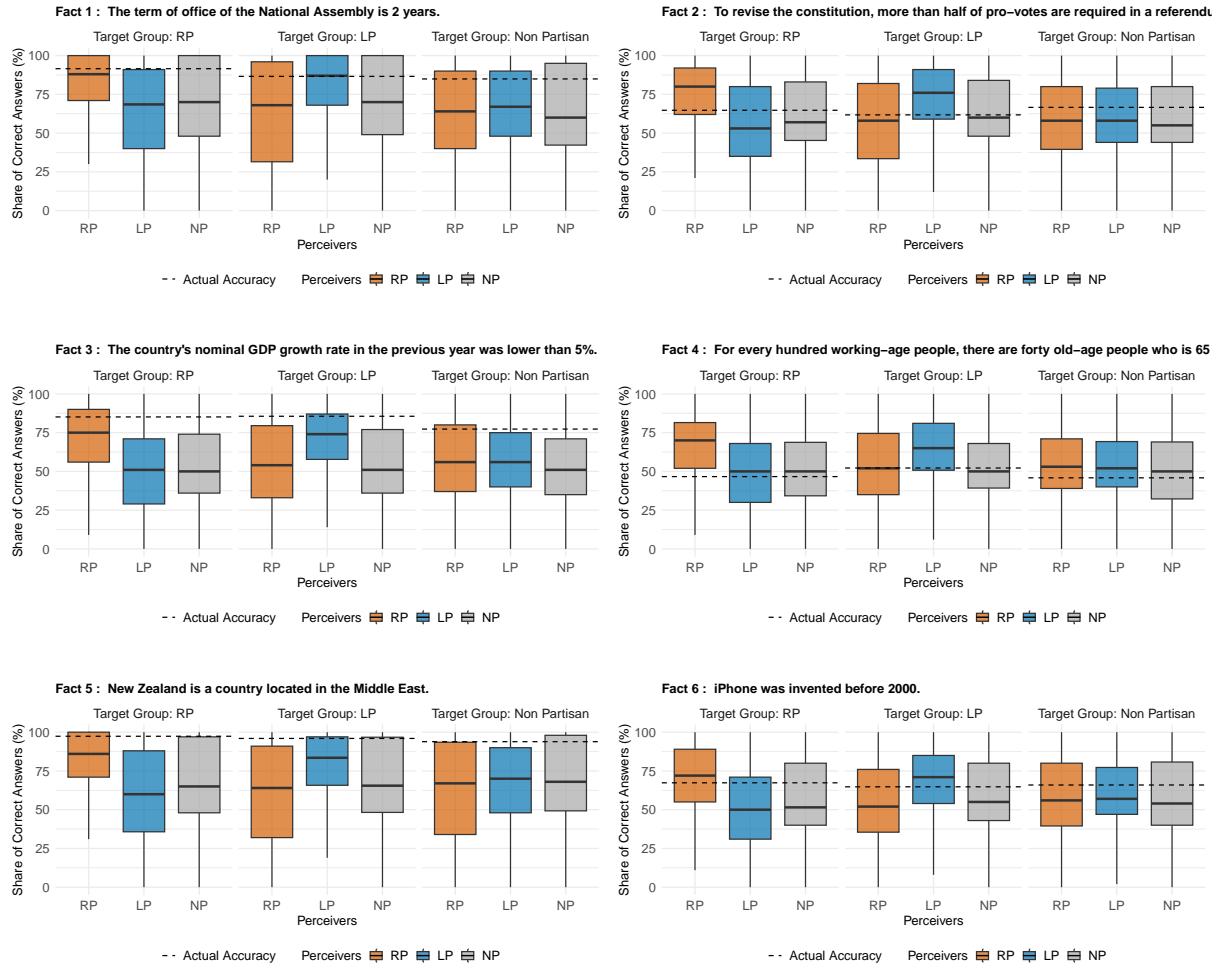
Figure A7 presents the results. Contrary to the case of non-partisan facts, even the actual accuracy rates differ substantially across political groups.<sup>18</sup> That said, we observe a pattern similar to partisan disbelief for non-partisan facts. However, the degree of partisan bias in conspiracy theories is considerably larger than that in partisan disbelief for non-partisan facts. For example, the median supporter of party R believes that they are more knowledgeable about right-wing conspiracy theories than party L supporters by approximately 40 percentage points.

<sup>18</sup>We define “False” as the correct answer for all conspiracy theories, although there may be debate over whether a given conspiracy theory is indeed false.

## C Baseline Survey: Partisan Disbelief for Each Issue

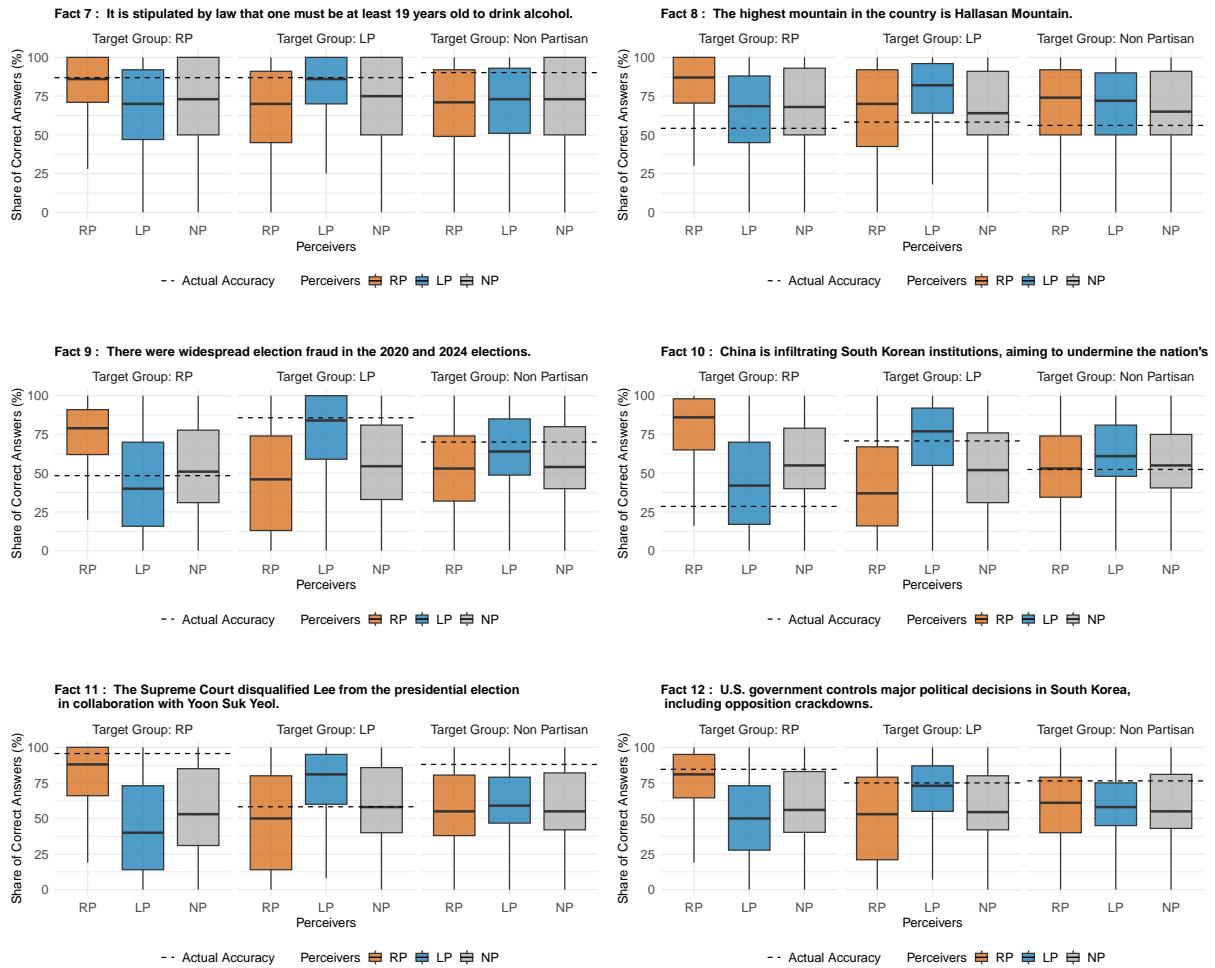
In the main text, we showed partisan disbelief for all questions combined. In this section, we study the partisan disbelief for each judgment task. Figure A8 reports target-based partisan belief for each non-partisan fact, and Figure A9 reports that for each conspiracy theory in South Korea. Figure A10 reports target-based partisan belief for each non-partisan fact, and Figure A11 reports that for each conspiracy theory in the U.S.. The results are consistent with the findings based on the average across tasks.

Figure A8: Partisan Disbelief for Each Question: South Korea



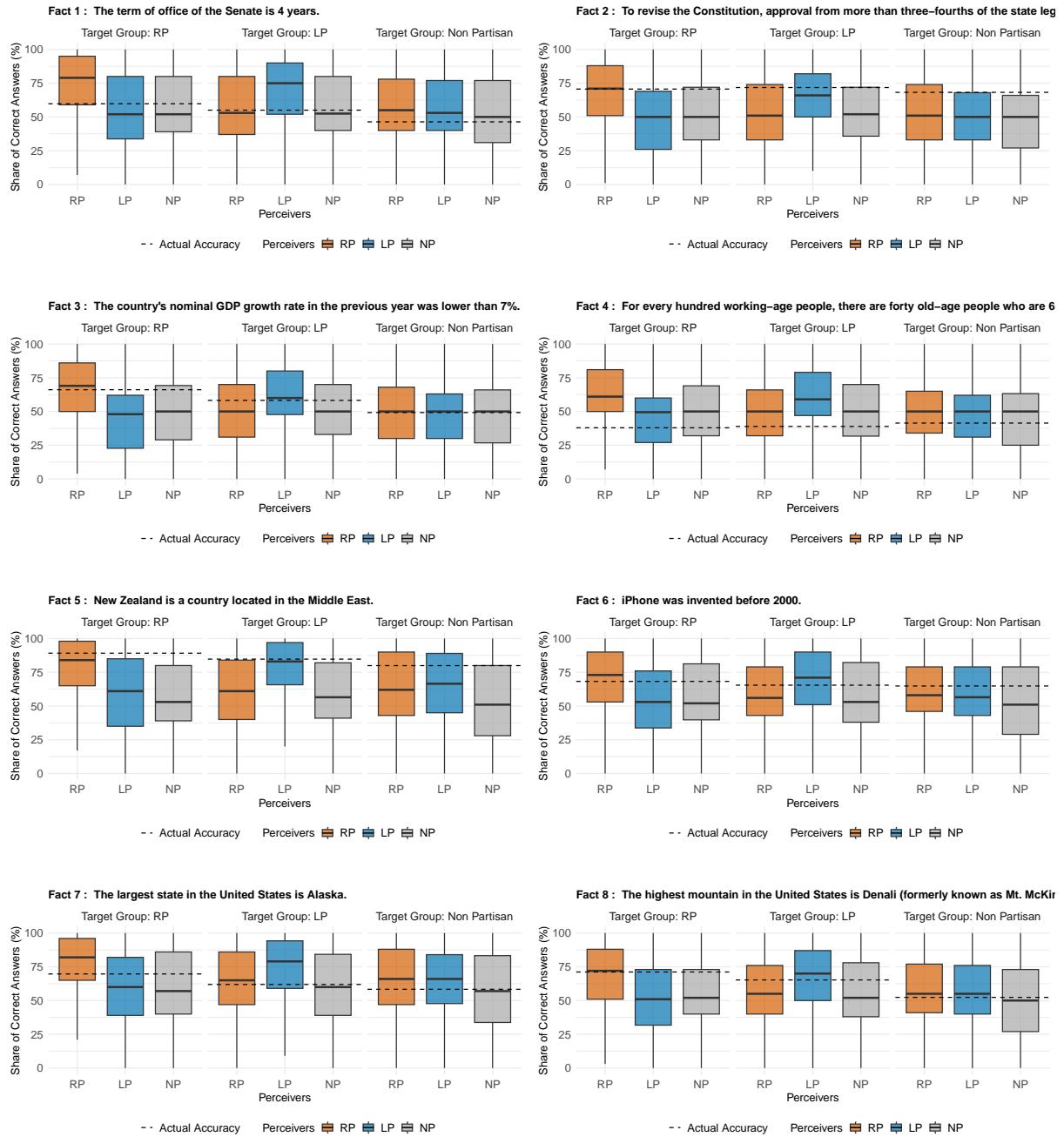
*Note:* These figures display the distributions of perceived accuracy for each question by perceiver and target groups in South Korea. Each panel corresponds to one statement. The boxes show the interquartile range (25th–75th percentiles) with medians indicated by black lines. Dashed lines represent the actual accuracy rates for each fact. Colors denote perceiver groups: blue for right-party supporters (RP), orange for left-party supporters (LP), and gray for non-partisans (NP).

Figure A9: Partisan Disbelief for Each Question: South Korea; Continued



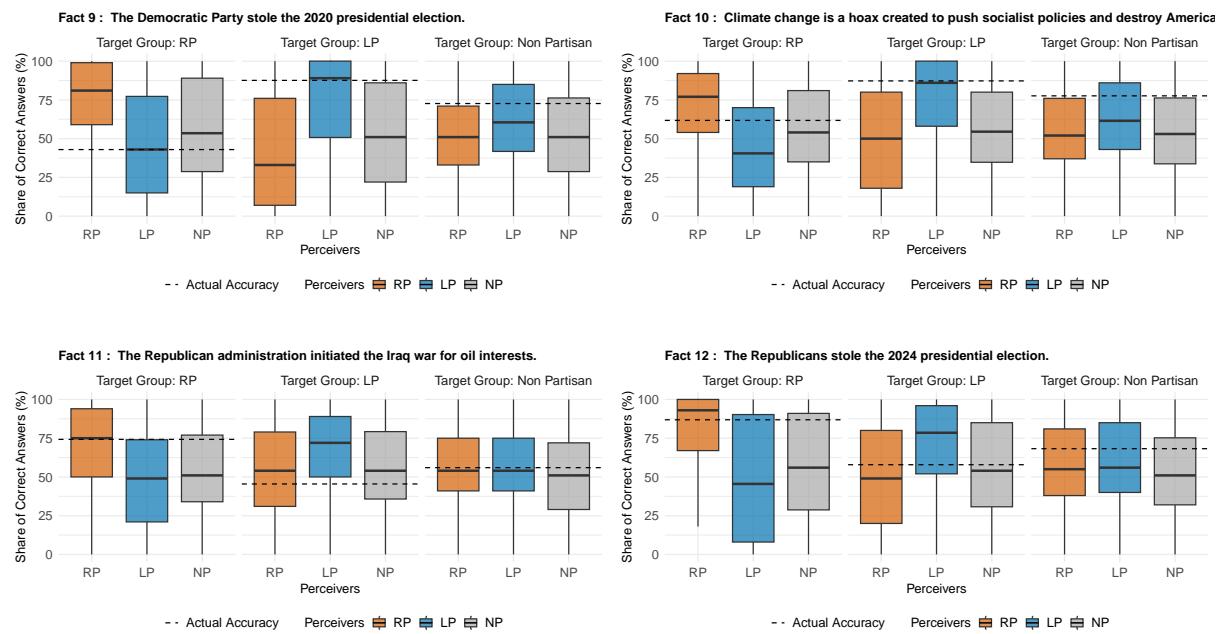
*Note:* These figures display the distributions of perceived accuracy for each question by perceiver and target groups in South Korea. Each panel corresponds to one statement. The boxes show the interquartile range (25th–75th percentiles) with medians indicated by black lines. Dashed lines represent the actual accuracy rates for each fact. Colors denote perceiver groups: blue for right-party supporters (RP), orange for left-party supporters (LP), and gray for non-partisans (NP).

Figure A10: Partisan Disbelief for Each Question: United States, Factual



*Note:* These figures display the distributions of perceived accuracy for each question by perceiver and target groups in the U.S.. Each panel corresponds to one statement. The boxes show the interquartile range (25th–75th percentiles) with medians indicated by black lines. Dashed lines represent the actual accuracy rates for each fact. Colors denote perceiver groups: blue for right-party supporters (RP), orange for left-party supporters (LP), and gray for non-partisans (NP).

Figure A11: Partisan Disbelief for Each Question: United States, Conspiracy



*Note:* These figures display the distributions of perceived accuracy for each question by perceiver and target groups in the U.S.. Each panel corresponds to one statement. The boxes show the interquartile range (25th–75th percentiles) with medians indicated by black lines. Dashed lines represent the actual accuracy rates for each fact. Colors denote perceiver groups: blue for right-party supporters (RP), orange for left-party supporters (LP), and gray for non-partisans (NP).

## D Experiment: Conspiracy Theory Questions

### D.1 Overview

After judgment tasks on non-partisan facts, respondents were asked to do another two tasks ( $j = 6, 7$ ). The tasks are almost the same as in the above, but tasks  $j = 6, 7$  were about whether a conspiracy theory is true or false. We presented the position most supported by the majority of R/L supporters. The list of statements used in these tasks can be seen in Table A3.

Table A3: Full List of True-or-False Questions in Experiment

		South Korea	United States
Pre-treatment	non-partisan facts	The term of the National Assembly is 2 years. (F)	To revise the Constitution, approval from more than three-fourths of the state legislatures is required. (T)
		New Zealand is a country located in the Middle East. (F)	New Zealand is a country located in the Middle East. (F)
Post-treatment	non-partisan facts	To revise the Constitution, a majority of votes in a national referendum is required. (T)	The country's nominal GDP growth rate in the last year was lower than 7%. (T)
		The iPhone was invented before 2000. (F)	The iPhone was invented before 2000. (F)
		The highest mountain in the country is Hallasan. (T)	The term of office in the Senate is 4 years. (F)
conspiracy theories		The Supreme Court colluded with Yoon Suk-yeol and decided to disqualify Lee Jae-myung from the presidential election.	The Republican administration initiated the Iraq War for oil.
		There was widespread election fraud in the parliamentary elections of 2020 and 2022.	The Democratic Party stole the 2020 presidential election.
non-partisan facts for education signals		The country's nominal GDP growth rate in the last year was lower than 5%. (T)	Alaska is the largest state in the United States. (T)
		By law, you must be at least 19 years old to drink alcohol. (T)	The highest mountain in the United States is Mt. McKinley (also known as Denali). (T)

### D.2 Treatment Effects on Disbelief

We first examine whether the treatment reduces disbelief for conspiracy theory questions. Table A4 shows the results. For South Korea, the point estimate is -0.028 with the standard error of 0.015. This is small, compared to the baseline disbelief of 0.343. For the US, the magnitude is larger. The treatment reduces the disbelief in conspiracy theory questions by 6.6 pt where the baseline is 23.9%.

Table A4: Treatment Effects on Partisan Disbelief: Conspiracy Theory

	(1) SK	(2) US
Treatment	-0.028 (0.015)	-0.066 (0.013)
Observations	2305	2792
Mean of outcome	0.343	0.239

Note: This table reports the estimated treatment effects on post-treatment partisan disbelief in South Korea (SK) and the United States (U.S.). The dependent variable is the average difference in perceived accuracy between in-group and out-group supporters across four conspiracy statements. Each coefficient represents the effect of being assigned to the treatment group ( $T_i = 1$ ) relative to the control group. Standard errors, shown in parentheses, are robust to heteroskedasticity.

### D.3 Information Processing

**Measurement** 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  $y_{i,j}^C$  as follows.

$$y_{i,j}^C = \begin{cases} y_{i,j} & \text{if } g(i) = g(j) \\ 1 - y_{i,j} & \text{if } g(i) \neq g(j) \end{cases} \quad (12)$$

Thus,  $y_{i,j}^C > 0$  means that respondent  $i$  updates the opinion toward that held by a majority of the opposing-party members.

**Specification** We run the following regression for the treated and control groups separately:

$$y_{i,j}^C = \beta^I \mathbb{1}\{s_{i,j} = I\} + \beta^O \mathbb{1}\{s_{i,j} = O\} + \eta_j + \varepsilon_{i,j}. \quad (13)$$

Given this specification, we hypothesize as follows:

- Signals affect information processing. That is,  $\widehat{\beta}^I < 0$  and  $\widehat{\beta}^O > 0$  for both the control and the treated groups. The higher value means
- 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.
- 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.

**Results** Table A5 presents the results. The starting point is the hypothesis that  $\hat{\beta}^I < 0$  and  $\hat{\beta}^O > 0$ . That is, in-group (resp. out-group) signals are expected to move beliefs away from (resp. toward) the opinion held by a majority of out-party members. Although this may appear straightforward at first glance, the results show that this pattern does not necessarily hold. For example, in the case of the binary opinion,  $\hat{\beta}^O < 0$  in South Korea, and  $\hat{\beta}^I > 0$  in the U.S.. Accordingly, the two subsequent hypotheses are also not supported.

Conspiracy theories are highly partisan issues; thus, respondents may infer the opposite meaning from the signals. For example, when a majority of the opposing party rejects a conspiracy theory, respondents may interpret this as evidence that the conspiracy is true. This could be a reason why the expected results were not obtained.

Table A5: Information Processing for Conspiracy Theory

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	SK	SK	SK	SK	US	US	US	US
	Dum	Dum	Cont	Cont	Dum	Dum	Cont	Cont
	C	T	C	T	C	T	C	T
In-Group Signal	-0.085 (0.046)	-0.154 (0.055)	0.125 (0.121)	0.195 (0.120)	0.345 (0.066)	0.205 (0.066)	1.180 (0.114)	0.966 (0.115)
Out-Group Signal	-0.085 (0.046)	-0.173 (0.055)	0.129 (0.121)	0.163 (0.120)	0.371 (0.066)	0.230 (0.065)	1.208 (0.114)	0.983 (0.114)
Observations	2278	2332	2278	2332	2814	2770	2814	2770

Note: This table reports the results of regressions examining information processing in conspiracy theory tasks for South Korea (SK) and the United States (U.S.). The dependent variable is an indicator for whether a respondent's post-signal belief aligns with the position held by a majority of out-party members. The key regressors are indicators for in-group and out-group signals, estimated separately for the control (C) and treated (T) groups. Columns labeled "Dum" refer to binary belief updates. Task fixed effects are included, and robust standard errors are shown in parentheses.

**Alternative measure** As an alternative measure, we redefine  $y_{ij} = J_{ij1} - J_{ij0} \in \{-1, 0, 1\}$  and redefine

$$y_{i,j}^C = \begin{cases} y_{i,j} & \text{if } g(i) = g(j) \\ -y_{i,j} & \text{if } g(i) \neq g(j). \end{cases} \quad (14)$$

This measure enables us to account for not only updates toward the correct decision ( $J_{ij0} = 0$  and  $J_{ij1} = 1$ ) but also updates toward the wrong decision ( $J_{ij0} = 1$  and  $J_{ij1} = 0$ ).

Table A6 reports the results when this alternative measure is adopted. The results diverge from our initial hypotheses even under this alternative measure.

Table A6: Information Processing for Conspiracy Theory: Alternative Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	SK	SK	SK	SK	US	US	US	US
	Dum	Dum	Cont	Cont	Dum	Dum	Cont	Cont
	C	T	C	T	C	T	C	T
In-Group Signal	-0.365 (0.081)	-0.461 (0.084)	-0.751 (0.242)	-0.611 (0.240)	0.041 (0.101)	-0.113 (0.104)	1.359 (0.228)	0.932 (0.229)
Out-Group Signal	-0.359 (0.081)	-0.485 (0.084)	-0.741 (0.242)	-0.674 (0.241)	0.070 (0.101)	-0.093 (0.104)	1.416 (0.227)	0.966 (0.228)
Observations	2278	2332	2278	2332	2814	2770	2814	2770

Note: This table reports the results of regressions analyzing information processing in conspiracy theory tasks using an alternative measure of belief updating. The dependent variable takes values in  $-1, 0, 1$  to capture both correct and incorrect updates, where positive values indicate updates toward the correct position and negative values indicate updates toward the wrong position. The key regressors are indicators for in-group and out-group signals, estimated separately for the control (C) and treated (T) groups. Columns labeled "Dum" correspond to binary belief updates. Task fixed effects are included, and robust standard errors are shown in parentheses.

## E Experiment: Education Signals

At the end of the survey, respondents were asked to do another two tasks ( $j = 8, 9$ ). The tasks are almost the same as in the tasks for partisan signals about non-partisan facts, but at this time, each respondent randomly received one of the following two signals: the signal telling them the majority of college graduates' opinion or the signal telling them the majority of non-college graduates' opinion. The list of statements used in these tasks can be seen in Table A3.

Figure A12: Education Signal

Your judgement on the previous page is that "Alaska is the largest state in the United States" is TRUE.

According to our previous survey, a majority of college graduates also say that "Alaska is the largest state in the United States" is TRUE.

**A majority of college graduates**

say that

"Alaska is the largest state in the United States" is TRUE.



Please choose the appropriate sentence based on the above. You cannot move to the next question without choosing the correct answer.

- My initial judgment was different from the judgment by a majority of college graduates.
- My initial judgment was the same as the judgment by a majority of college graduates.

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** The estimated effect of in-group signals depends on the accuracy rate of the judgment tasks before receiving the signals. For instance, if everyone correctly judges whether statements are true or false even before receiving any signals, they will not respond to those signals (thus, the estimated effect will be small). Because the accuracy rates differ between the tasks used for partisan signals and those used for education signals, this poses a challenge when comparing the effects of the two types of signals. To address this issue, we restrict samples  $(i, j)$  to the control

group whose pre-signal answers are wrong. We run the following three regressions. For  $j = 3, 4, 5$ , we run

$$y_{i,j} = \tilde{\beta}_1 \mathbb{1}\{s_{i,j} = I\} + \text{const.} + \varepsilon_{i,j}. \quad (15)$$

For  $j = 8, 9$ , we run

$$y_{i,j} = \tilde{\beta}_2 \mathbb{1}\{s_{i,j} = \text{College}\} + \text{const.} + \varepsilon_{i,j} \quad (16)$$

$$y_{i,j} = \tilde{\beta}_3 \mathbb{1}\{s_{i,j} = I_E\} + \text{const.} + \varepsilon_{i,j} \quad (17)$$

(15) estimates the effect of in-group partisan signals, while (16) and (17) estimate the effects of education signals.

The effect of education signals may differ depending on whether the signal refers to the opinions of college graduates or non-graduates, for two reasons. First, regardless of respondents' own educational background, they may respond more strongly to signals from college graduates. Second, respondents may react more strongly to signals from individuals who share the same educational profile.  $\mathbb{1}\{s_{i,j} = \text{College}\}$  takes one if the signal comes from college-graduates, whereas  $\mathbb{1}\{s_{i,j} = I_E\}$  takes one if the signal comes from those who share the same education profile with respondent  $i$ . Therefore, equation (16) captures the first scenario, whereas equation (17) captures the second.<sup>19</sup>

By comparing  $\hat{\beta}_1$  with  $\hat{\beta}_2$  and  $\hat{\beta}_3$ , 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.

**Results** Table A7 shows the results in South Korea, and Table A8 shows the results in the U.S.. Specifically, column (1) reports  $\hat{\beta}_2$ , (2) reports  $\hat{\beta}_3$ , and (3) reports  $\hat{\beta}_1$  for the binary opinion. Columns (4)–(6) present the corresponding estimates for the continuous opinion measure.

The results indicate that respondents do not respond significantly to education signals. For instance, in South Korea,  $\hat{\beta}_2$  is positive and statistically significant; however, respondents react to signals from college graduates only 7 percentage points more than to those from non-graduates. This effect is substantially smaller than that of in-group partisan signals (17.3 percentage points).

These results yield two key implications. First, partisan in-group bias in information processing is substantial. Second, the findings suggest that this bias cannot be simply attributed to misperceptions about the educational composition of each political group. If partisan disbelief merely reflected such misperceptions, the effect of partisan signals should not exceed that of education signals.

---

<sup>19</sup>In the pre-analysis plan, we specified (16), but (17) was not induced.

Table A7: In-Group Bias Compared to Education Signals: South Korea

	(1) Dummy Educ	(2) Dummy Educ	(3) Dummy Party	(4) Continuous Educ	(5) Continuous Educ	(6) Continuous Party
In-Group Signal		-0.004 (0.006)	0.173 (0.016)		-0.024 (0.001)	0.140 (0.021)
College Signal	0.007 (0.001)			0.000 (0.022)		
Observations	1185	1185	1133	1185	1185	1133
Mean of outcome	0.070	0.070	0.309	0.257	0.257	0.530

This table compares the magnitude of in-group bias in information processing with the effect of education-based signals in South Korea. The dependent variable measures the change in respondents' opinions before and after receiving a signal, conditional on incorrect pre-signal answers. The key regressors are indicators for in-group partisan signals and education signals. Columns (1)–(3) report results for the binary ("dummy") update measure, while columns (4)–(6) use the continuous update measure. Task fixed effects are included, and robust standard errors are shown in parentheses. "College Signal" indicates signals from college graduates, and "In-Group Signal" refers to signals from respondents' own partisan group.

Table A8: In-Group Bias Compared to Education Signals: United States

	(1) Dummy Educ	(2) Dummy Educ	(3) Dummy Party	(4) Continuous Educ	(5) Continuous Educ	(6) Continuous Party
In-Group Signal		-0.038 (0.038)	0.021 (0.021)		-0.038 (0.014)	0.090 (0.017)
College Signal	0.032 (0.027)			0.057 (0.030)		
Observations	827	827	1544	827	827	1544
Mean of outcome	0.522	0.522	0.466	0.606	0.606	0.532

This table compares the magnitude of in-group bias in information processing with the effect of education-based signals in the U.S.. The dependent variable measures the change in respondents' opinions before and after receiving a signal, conditional on incorrect pre-signal answers. The key regressors are indicators for in-group partisan signals and education signals. Columns (1)–(3) report results for the binary ("dummy") update measure, while columns (4)–(6) use the continuous update measure. Task fixed effects are included, and robust standard errors are shown in parentheses. "College Signal" indicates signals from college graduates, and "In-Group Signal" refers to signals from respondents' own partisan group.

## F Experiment: Experimenter Demand Effect

While previous research has shown that experimenter demand effects are limited in online survey experiments (Mummolo and Peterson, 2019), it remains important to examine whether our findings could be driven by such effects. To ensure that our results are not driven by experimenter demand, we reanalyze the data after excluding respondents who appeared to pander to the 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.<sup>20</sup>

We asked 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 (Mummolo and Peterson, 2019). Thus, such respondents are subject to the experimenter demand effect.

We excluded such respondents, which left us 2112 respondents in South Korea and 2495 respondents in the U.S..

**Results** Table A9 presents the treatment effect on partisan disbelief; Table A10 reports the in-group bias in information processing in the control group; Table A11 presents the treatment effect on this in-group bias; and Table A12 shows the treatment effect on affective polarization. Overall, the results are consistent with those from the main analysis, indicating that experimenter demand effects are not a serious concern in our experiment.

---

<sup>20</sup>The same approach was taken by Kishishita and Matsumoto (2024). A similar approach was also taken by Dhar et al. (2022). They measured each respondent’s propensity to provide socially desirable answers based on a questionnaire used in social psychology. They found that the treatment effect did not differ across propensities.

Table A9: Treatment Effects on Partisan Disbelief

	(1)	(2)
	SK	US
Treatment	-0.049 (0.011)	-0.075 (0.010)
Observations	2112	2495
Mean of outcome	0.227	0.191

This table reports the estimated treatment effects on partisan disbelief after excluding respondents identified as susceptible to experimenter demand effects. The dependent variable is the post-treatment measure of partisan disbelief, defined as the difference in perceived accuracy between in-group and out-group supporters across non-partisan factual questions. Each coefficient represents the effect of being assigned to the treatment group ( $T_i = 1$ ) relative to the control group. Standard errors, shown in parentheses, are robust to heteroskedasticity.

Table A10: In-Group Bias in Information Processing (Control Group)

	(1)	(2)	(3)	(4)
	SK	SK	US	US
	Dummy	Continuous	Dummy	Continuous
In-Group Signal	0.059 (0.010)	0.082 (0.007)	0.008 (0.010)	0.048 (0.012)
Observations	3135	3135	3759	3759
Mean of outcome	0.100	0.424	0.174	0.486

Note: This table reports the estimated in-group bias in information processing for the control group after excluding respondents susceptible to experimenter demand effects. The dependent variable measures the change in respondents' judgments before and after receiving a signal, defined either as a binary ("dummy") update or a continuous update. The key regressor is an indicator for whether the signal originates from in-group members. Task fixed effects are included, and robust standard errors are reported in parentheses.

Table A11: Treatment Effects on In-Group Bias in Information Processing

	(1)	(2)	(3)	(4)
	SK	SK	US	US
	Dummy	Continuous	Dummy	Continuous
In-Group Signal	0.059 (0.010)	0.082 (0.009)	0.008 (0.011)	0.048 (0.012)
Treatment	0.017 (0.008)	0.023 (0.018)	-0.025 (0.011)	0.022 (0.021)
In-Group Signal x Treatment	-0.035 (0.002)	-0.063 (0.033)	0.014 (0.009)	-0.047 (0.022)
Observations	6336	6336	7485	7485
Mean of outcome	0.100	0.420	0.165	0.485

Note: This table reports the treatment effects on in-group bias in information processing after excluding respondents susceptible to experimenter demand effects. The dependent variable measures the change in respondents' judgments before and after receiving a signal, defined either as a binary ("dummy") update or a continuous update. The key variable *In-Group Signal* × *Treatment* captures the differential effect of in-group versus out-group signals for treated respondents relative to the control group. Task fixed effects are included, and robust standard errors are shown in parentheses.

Table A12: Treatment Effects on Affective Polarization

	(1) SK Unfav	(2) SK Uncomf	(3) US Unfav	(4) US Uncomf
Treatment	-0.027 (0.013)	-0.001 (0.013)	-0.088 (0.013)	-0.026 (0.017)
Observations	2112	2112	2495	2495
Mean of outcome	0.530	0.439	0.497	0.276

Note: This table reports the estimated treatment effects on affective polarization after excluding respondents identified as susceptible to experimenter demand effects. The dependent variable is the difference in favorable feelings toward in-party versus out-party members, measured either by self-reported warmth ratings (*Unfav*) or by (un)comfortableness in social relationships with out-party members (*Uncomf*). Each coefficient represents the effect of being assigned to the treatment group ( $T_i = 1$ ) relative to the control group. Robust standard errors are shown in parentheses.

## **G Questionnaires**

### **G.1 Baseline Survey in South Korea**

[Link to PDF \(in Korean\)](#)

### **G.2 Baseline Survey in the U.S.**

[Link to PDF](#)

### **G.3 Experiment in South Korea**

[Link to PDF \(in Korean\)](#)

### **G.4 Experiment in the U.S.**

[Link to PDF](#)