The International Landscape of Peace: How Third-Party Interactions Shape Public Opinion on Rapprochement

Jungmin Han*

January 9, 2026

Please read the latest version of the paper and the appendix.

Abstract

Under what international contexts are citizens more likely to support peace with a foreign adversary? To explain public opinion on interstate peace, existing research has largely focused on factors within rivalry dyads, overlooking the influence of a broader international environment in which rivalries are situated. This study addresses this gap by introducing a novel model that explains how international interactions involving third-party states shape public perceptions of state actors. Model simulations suggest that interstate conflict involving third-party actors—such as between an ally and a rival or between an ally and a rival's ally—can heighten perceptions of inter-group confrontation, reinforcing the adversary's enemy image and increasing opposition to conciliatory policies. The theoretical claims are tested using both experimental and observational methods. The first study employs a survey experiment to assess how exposure to a news video clip of conflict or cooperation between allied and enemy states (e.g., China-Japan, North-South Korea) influences American attitudes toward China. The second study applies supervised machine learning to over one million tweets to examine how fluctuations in U.S.–North Korea relations between 2017 and 2018 shaped public sentiment in South Korea and Japan toward their respective rivals, North Korea and China. The findings of both studies indicate that exposure to third-party conflicts increases the likelihood that individuals perceive rival states more negatively and show stronger resistance to rapprochement. These results demonstrate that public opinion on interstate peace is both socially embedded, shaped not only by bilateral dynamics but also by broader patterns of interstate behavior.

*Postdoctoral Researcher, Department of Political Science, Trinity College Dublin (jhan@tcd.ie).

1. Introduction

South Korean presidents Roh Moo-hyun and Moon Jae-in—close political allies with shared personal and professional backgrounds—pursued notably similar foreign policy strategies toward North Korea. Both were affiliated with the same liberal party, championed democratic reform, and began their careers as human rights lawyers who gained prominence through grassroots activism (BBC News, 2017). Consistent with these shared commitments, their administrations prioritized peaceful engagement with Pyongyang, achieving diplomatic milestones through high-level inter-Korean summits. President Roh's conciliatory policy culminated in a historic 2007 meeting with Kim Jong-il to discuss peace and cooperation (Onishi, 2007). More than a decade later, President Moon adopted a similarly pacific approach, initiating a peace initiative that led to three summits with Kim Jong-un in 2018 (Botto & Jo, 2018). In both cases, the summits produced pledges from North Korea on denuclearization and military de-escalation, signaling a renewed era of inter-Korean dialogue.

Yet despite these parallel diplomatic efforts and comparable concessions from North Korea, public reactions diverged sharply. Roh's dovish approach faced persistent skepticism: a majority of South Koreans disapproved of his foreign policy as early as 2004 (53.6% disapproval) (East Asia Institute, 2004), and by the time of the 2007 summit, support for his North Korea policy remained low—only 10.9% in approval, versus 47.8% disapproval (The Hankyoreh, 2007). In contrast, Moon's 2018 engagement strategy was widely endorsed. A majority of South Koreans (64.4%) expressed support for his North Korea policy (Jung et al., 2018, 100), accompanied by a surge in public trust toward North Korea and a marked decline in perceptions of the North as the primary national threat.[1] Survey data from Gallup (2018) further reveal that Moon's initiative enjoyed cross-party support, including 98% approval from Democratic Party supporters and substantial backing from opposition party members (67% of Bareunmirae and 50% of Liberty Korea Party affiliates). These contrasting public responses likely influenced the sustained momentum of Moon's initiative and the early collapse of Roh's.[2]

[1] Trust in North Korea increased from 28.1% to 54.7% between 2017 and 2018, while the percentage of South Koreans identifying the North as the greatest threat fell from 63.7% to 32.8% (Jung et al., 2018, 72–138).

[2] Public support is vital for rapprochement because resistance from hawkish elites and their supporters can

The stark divergence in South Korean public responses to these two comparable peace initiatives poses a puzzle for existing studies on public opinion toward interstate peace. One prominent line of scholarship emphasizes the role of behavioral dynamics between rival states. Cooperative behaviors based on reciprocity (Colaresi, 2004; Kupchan, 2012; Davies & Johns, 2016) and costly signals (Albuyeh & Paradis, 2018; Kertzer et al., 2020; Kim, 2021; Fu & Lee, 2024) are highlighted as a means to achieve public approval for peace. However, in both administrations, North Korea made comparable overtures—including summit diplomacy and pledges of denuclearization—yet only Moon's initiative received broad public approval. If reciprocity or signal cost were the primary drivers, similar public reactions would be expected across both periods. Another body of literature underscores the importance of foreign policy reputation, suggesting that hawkish leaders or parties are more credible in advocating peace—a logic often referred to as the "Only Nixon goes to China" effect (Schultz, 2005; Kreps et al., 2018; Mattes & Weeks, 2019). Yet this framework also falls short in explaining the South Korean case. Both Roh and Moon were members of the same liberal party and were widely viewed as dovish leaders, yet public responses to their peace efforts diverged significantly.

The difficulty existing approaches face in explaining the divergent public responses to Presidents Roh and Moon's similar peace initiatives may stem, in part, from their limited analytical scope. By focusing primarily on dynamics within the rivalry dyad—such as bilateral signaling and leader reputation—these studies overlook the broader international environment in which rivalries are situated. International rivalries rarely develop or persist in isolation; they are embedded within intricate networks of alliances, enmities, and historically contingent relationships involving multiple state actors. Historical cases such as the U.S.–China rapprochement in the 1970s, which influenced Soviet foreign policy and facilitated the normalization of China–Japan relations (Herrmann, 1985; Goldstein & Freeman, 1990), or NATO's shifting posture in response to Serbian aggression during the Kosovo conflict (Goldstein & Pevehouse, 1997), highlight how third-party behaviors can reshape the broader landscape of international conflict. Recent empirical studies further underscore

threaten the process by pressuring leaders to abandon peace efforts to protect their political standing (Vasquez, 1993; Colaresi, 2004). Domestic opposition can also weaken the credibility of cooperative signals to the adversary, undermining the chances for lasting peace (Brandt et al., 2008; Rasler et al., 2013).

that such interdependence among state actors is a broad feature of international conflict and cooperation (Rider & Owsiak, 2021; Lee, 2022; Edgerton, 2024; Beardsley, 2024).

This study addresses this gap by examining how individuals revise their perceptions of foreign adversaries within a broader international context. It introduces a model of national image updating grounded in the observation of interstate behaviors, particularly those involving third-party states. Building on insights from constructivists and social psychologists, the model assumes that individuals' mental representations of foreign states—whether perceived as allies or enemies—are not fixed, but socially constructed and responsive to changing international environments. Importantly, the model moves beyond traditional dyadic frameworks by incorporating third-party interactions, positing that individuals form national images not in isolation, but in relation to a wider network of alliances and rivalries. Relevantly, two specific mechanisms of social influence are highlighted in the model: Direct Third-Party Interaction (DTI), where a target state interacts with a third-party state, and Peripheral Third-Party Interaction (PTI), where third-party actors engage with each other but not directly with the target. The model's simulation results demonstrate that increased hostility in these third-party interactions amplifies perceptions of intergroup antagonism, solidifies enemy images, and weakens support for reconciliation efforts.

To test the theoretical claims, this study employs a multi-method approach. Study 1 tests the causal impact of third-party interactions on public attitudes toward a foreign adversary by conducting a pre-registered survey experiment on how Americans view rapprochement toward China. Participants were randomly assigned to watch one of four news video clips depicting either conflict or cooperation between two states aligned with opposing blocs—Japan and China (DTI), or North and South Korea (PTI). The results provide supportive evidence for the theoretical expectations. Respondents who observed hostile interactions between Japan and China rated China more negatively and were significantly less supportive of cooperative U.S. foreign policy toward China. Moreover, this shift was accompanied by a polarization of national images: enemy images (e.g., North Korea) became more pronounced, while ally images (e.g., Japan, South Korea) also intensified. The effects of PTI—interactions between South and North Korea—mirrored this pattern but were generally weaker.

To complement the experimental evidence with observational data, Study 2 leverages

over one million tweets from South Korea and Japan to examine how real-world fluctuations in U.S.–North Korea relations between 2017 and 2018 shaped public sentiment toward rival states. Using supervised machine learning models, this study estimates the prevalence of hostile attitudes (enemy image) and opposition to conciliatory policies (anti-dovish sentiment) in tweets referencing North Korea and China. The findings from South Korea are consistent with the theory: during periods of heightened U.S.–DPRK tension (DTI), anti-North Korean sentiment and domestic criticism of pro-peace actors increased significantly. Similar, albeit weaker, patterns emerged in Japan, where conflict between the U.S. and North Korea (PTI) corresponded with increased hostile portrayals of China and greater criticism of pro-China political actors.

Beyond its primary contribution of identifying the role of third-party states in shaping public opinion on rapprochement, this study advances two foundational areas of scholarship in international politics and public opinion. First, this study contributes to the microfoundations of social theory in international relations by bridging constructivist insights with individual-level perceptions. While constructivist scholars have long emphasized the role of norms, identity, and social structures in shaping international outcomes (Katzenstein, 1996; Wendt, 1999; Kriesberg, 2007), they have paid less attention to how individuals internalize and update these social meanings. This research extends that tradition by showing how observable interstate behaviors shape individuals' mental representations of the international system and the social status of foreign states within it. Second, this research engages with longstanding debates on the structure and stability of public opinion. While early studies portrayed mass attitudes on foreign affairs as unstructured and inconsistent (Converse, 1964; Zaller, 1992), more recent scholarship has uncovered systematic patterns in how individuals form and revise foreign policy preferences (Aldrich et al., 2006; Berinsky, 2009; Kertzer & Zeitzoff, 2017). By showing that citizens respond not only to direct interactions with rival states but also to the broader configuration of international relations, this paper offers further evidence that public opinion on foreign policy is meaningfully structured and responsive to changes in the global environment.

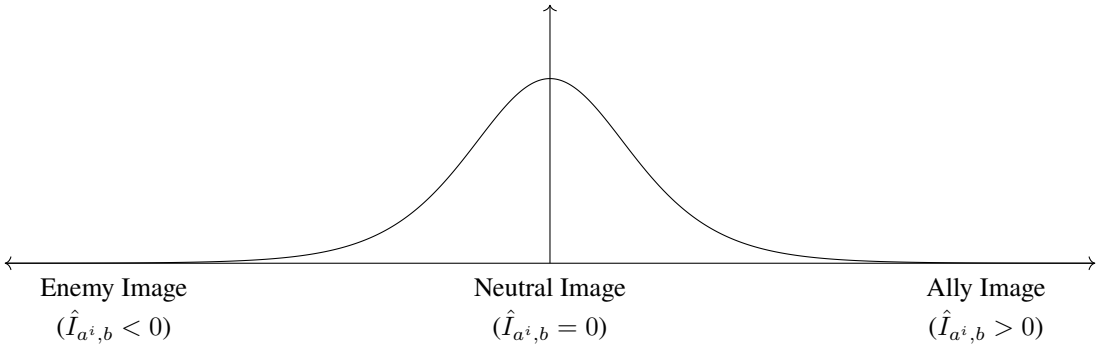
2. Mapping National Images Amid International Rivalry

The theoretical argument of this paper is grounded in the study of national image. Individuals do not use all available information when shaping their foreign policy attitudes. Instead, when thinking about foreign affairs, the majority of people tend to employ a national image as a heuristic cue "to simplify a complex international environment and to guide perceptions of, and responses toward, particular countries" (Hurwitz & Peffley, 1990, 6). This mental label, once established, provides an important cognitive framework that structures how people understand the attributes and behaviors of foreign nations (Jervis, 1976; Herrmann & Fischerkeller, 1995; Alexander et al., 2005). Public attitudes toward reconciliation with a foreign adversary are therefore likely shaped by deeply ingrained negative perceptions of that country.

Moreover, in the context of international rivalry, individuals tend to classify international actors based on how strongly they are perceived as allies or enemies. Image theorists argue that perceptions of foreign states are shaped by a combination of factors, including goal compatibility, relative power, and cultural status (Herrmann & Fischerkeller, 1995; Alexander et al., 2005). These factors contribute to forming various types of national images—such as ally, enemy, barbarian, imperialist, and dependent—each representing distinct cognitive frameworks for interpreting state behavior . Importantly, because foreign policy attitudes are often driven by concerns over national security and strategic interests (Cottam, 1977; Peffley & Hurwitz, 1992), individuals tend to simplify these categories by focusing on whether a given state is an ally or an enemy (Boulding, 1956; Herrmann, 1985; Kaplowitz, 1990; Schafer, 1997). This ally-enemy distinction becomes even more pronounced in contexts with the history of longstanding conflicts, where "us versus them" thinking is intensified (Bar-Tal & Halperin, 2013).

Yet, individuals do not necessarily interpret states through a strict ally-enemy dichotomy. Rather, they may situate them along a continuum that reflects varying intensities of perceived alliance or antagonism. Two primary dimensions shape these perceptions: (1) the extent to which a state is viewed as posing a threat (*threat perception*), and (2) the degree to which its political or ideological goals are seen as aligning with those of the individual's own nation (*goal compatibility*) (Herrmann & Fischerkeller, 1995; Alexander et al., 2005).

Figure 1: Distribution of National Images Across the Enemy-Ally Continuum



States seen as highly threatening or pursuing conflicting objectives are more likely to be viewed as enemies, while those perceived as less threatening and more aligned in goals are more likely to be seen as allies.

Furthermore, people may construct national images by situating a country within a broader network of relationships. Specifically, international rivalries often lead to the mental sorting of states into enemy and ally groupings. In the wake of internationalized wars or ongoing hostilities, individuals tend to minimize distinctions among allied states, while simultaneously intensifying the out-group perception of countries that fall into the opposing social category (Migdal et al., 2001; Risse-Kappen, 2016). This logic of social categorization may even apply to states not directly involved in the conflict. A general sense of in-group versus out-group affiliation can influence how people interpret the social position of third-party states by evaluating whether they aid or hinder their own nation's success in international competition (Bilali, 2010; Eicher et al., 2013). Accordingly, individuals assign national images to these states along the ally–enemy continuum based on perceived political or ideological characteristics that indicate alignment with—or opposition to—key rival or partner states.

Drawing from these assumptions, Figure 1 depicts how people amid international rivalry map the national images of multiple states they recognize in the world politics. Here, the x-axis of the graph represents the extent to which target states (b) are perceived by an individual i from a source state (a) as either an enemy or an ally ($\hat{I}_{a^ib}$), while its y-axis represents the density of states that correspond to these perceptions. The shape of the distribution approximates a logistic curve, which reflects the reality that only a small number of states are perceived as strong allies or adversaries—typically those involved in direct co-

operation or conflict—and are thus located at the outer ends of the continuum. In contrast, states not directly engaged in international rivalries are placed along the continuum based on their perceived political or ideological alignment with key rival or ally states. As a result, most countries tend to cluster near the center of the distribution, as few are assigned strong enemy or ally labels. Overall, despite variation in the intensity of national images, individuals tend to interpret the international system through a binary lens, framing global politics as a contest between opposing groups of allies and enemies.

3. Modeling National Image Updating Through Third-Party Interactions

While national images often serve as stable cognitive shortcuts, they are not impervious to change. Individuals can adjust their perceptions of foreign states in response to new information. In particular, international events—such as episodes of cooperation or conflict between state actors—can influence public attitudes by either reinforcing or challenging existing views about a state's alignment or hostility (Jervis, 1976; Peffley & Hurwitz, 1992). Building on this premise, this section introduces a model of national image updating based on the observation of interstate interactions. Crucially, these interactions need not involve one's own country directly; rather, perceptions may also shift based on how foreign states interact with each other. Constructivist views in international relations emphasize that national images are socially constructed and subject to change as individuals revise their understanding of the global landscape (Wendt, 1994; Katzenstein, 1996; Herrmann, 2017). In line with this view, the proposed model assumes that individuals' mental representations of the international system are fluid and responsive to changes in the broader pattern of interstate relations.

Equation 1 presents the model by which individuals update their national image of a foreign state. On the left-hand side of the equation, $\hat{I}_{a^i b,t}$ denotes the revised image that an individual i from country a holds of country b at time t. According to the right-hand side of the equation, this updated image results from a function combining individuals' prior images of state actors ($\hat{I}_{t-1}$) and the observed behaviors among them ($\hat{B}_{t-1}$). When an individual perceives n different states, $\hat{I}_{t-1}$ is an n-dimensional vector representing the

national images associated with each of those states at time t-1. On the other hand, $\hat{B}_{t-1}$ can be expressed as a $n \times n$ matrix that captures any perceived interactions between all pairs of countries, each associated with a particular national image at time t-1. Each element in this matrix reflects the perceived nature of the interaction between two countries—where higher positive values indicate more cooperative behavior, and more negative values indicate more hostile interactions. The interplay of these two components gives rise to the following four explanatory dimensions.

$$
\begin{aligned}
\hat{I}_{a^i b,t} = \quad & \hat{I}_{a^i b,t-1} && \text{: Prior National Image} \\
& + \hat{B}_{a^i b,t-1} && \text{: Bilateral Interaction (BI)} \\
& + \frac{\sum_{c \in T} \hat{B}_{bc,t-1} \hat{I}_{a^i c,t-1}}{(n-2)|\overline{I}|} && \text{: Direct Third-Party Interaction (DTI)} \\
& + \frac{\Big(\sum_{c \in T} \sum_{d \in T} \hat{B}_{cd,t-1} \hat{I}_{a^i c,t-1} \hat{I}_{a^i d,t-1} \Big) \hat{I}_{a^i b,t-1}}{\binom{n-2}{2} |\overline{I}|^3} && \text{: Peripheral Third-Party Interaction (PTI)}
\end{aligned}
\tag{1}
$$

The first explanatory factor, $\hat{I}_{a^i b,t-1}$, suggests that the updated image of a target state is shaped by its pre-existing position along the ally-enemy continuum. National images are often sticky, functioning as cognitive shortcuts that help individuals make sense of complex international environments (Jervis, 1976). This persistence is reinforced by research showing that once formed, such images are resistant to change, especially when they align with dominant narratives or ideological frames (Herrmann et al., 1999). Furthermore, from a social identity perspective (Tajfel, 1981), altering one's perception of a foreign state may threaten group-based self-concepts, particularly when that state is associated with an "enemy" or "ally" category. In such cases, the national image is not just a cognitive tool but a socially embedded construct, making it even less malleable. Thus, due to both cognitive efficiency and identity-protective motivations, the prior image of a state provides a foundational reference point regarding how individuals process new information and update their perceptions.

The second component of the model, $\hat{B}_{a^i b,t-1}$, suggests that individuals revise their image of a target state (b) in response to its bilateral interactions (BI) with their own state (a). This mechanism reflects the idea that foreign policy behavior—particularly acts of cooperation or hostility—serves as a key source of information through which individuals infer the intentions and character of other states. Recent research on the microfoundations of rapprochement has offered both direct and indirect evidence that increased cooperative signaling between international rivals is associated with lower levels of perceived hostility toward the foreign adversary (Davies & Johns, 2016; Kertzer et al., 2020; Mattes & Weeks, 2019, 2022; Fu & Lee, 2024). These studies highlight how citizens are responsive to the tone and substance of dyadic interstate behavior, often interpreting cooperation as a cue for trustworthiness and aggression as a sign of threat. Consistent with this empirical pattern, the model suggests that more cooperative bilateral interactions ($\hat{B}_{a^i b,t-1} > 0$) are associated with a more favorable, ally-like image of the target state, while hostile encounters ($\hat{B}_{a^i b,t-1} < 0$) strengthen its enemy image.

What sets this model apart from the existing views is its emphasis on the role of international behaviors that extend beyond direct bilateral relations. As discussed in the previous section, individuals situated within the context of international rivalry often perceive global politics as a competitive arena between opposing groups of allies and enemies. In this view, a target state's perceived image is not formed in isolation but is influenced by its broader social associations—particularly how it is aligned with other states that are categorized as either allies or adversaries. Accordingly, interstate interactions involving third-party states become crucial to the image updating process. This model identifies two distinct types of such interactions: (1) Direct Third-Party Interaction (DTI), where the target state engages directly with a third-party state, and (2) Peripheral Third-Party Interaction (PTI), where the target state is not a direct participant, but perceptions shift based on its associations with the third-party actors involved.

The first social component, DTI, reflects the degree to which a target state en-

gages in hostile or cooperative behavior toward third-party states that the observer perceives as allies or enemies. Structural balance theory, originally developed in the context of interpersonal relationships (Heider, 1946, 1982), posits that individuals seek cognitive consistency in their evaluations of social ties—preferring balanced relationships such as "the friend of my friend is my friend" or "the enemy of my friend is my enemy." Applied to international politics, this logic suggests that individuals evaluate foreign states not only based on direct bilateral relations but also in terms of how those states interact with others that are socially important to them. For instance, when a state perceived as an ally is targeted with aggression by a third-party state, observers are likely to adopt a more negative view of the aggressor, reinforcing its image as an adversary. Conversely, when a rival state cooperates with an ally, this challenges prior assumptions and creates psychological inconsistency. To restore a sense of balance, individuals may reassess the rival's image and downplay its adversarial character.

The model integrates this logic into its formal structure. Specifically, $\hat{B}_{bc,t-1}\hat{I}_{a^ic,t-1}$ represents how a target state (b) interacts with a third-party state (c), weighted by the image that individual i from country a holds of state c at time $t-1$. If a third-party state is viewed as an ally ($\hat{I}_{a^ic,t-1} > 0$), cooperative behavior between the target and the ally ($\hat{B}_{bc,t-1} > 0$) leads to a more favorable image of the target ($\hat{I}_{a^ib,t} > \hat{I}_{a^ib,t-1}$). Conversely, hostile interactions between a target and the ally ($\hat{B}_{bc,t-1} < 0$) increase the perceived negative image of the target state ($\hat{I}_{a^ib,t} < \hat{I}_{a^ib,t-1}$). Additionally, the overall effect depends on both the intensity of the interaction ($\hat{B}_{bc,t-1}$) and the strength of the observer's prior image of the third-party state ($\hat{I}_{a^ic,t-1}$). Recognizing that a target state may interact with multiple third-party states (T), the model aggregates these effects to capture the broader influence of the target's direct social interactions on its image. Also, to account for the number of third-party states and the scaling of their image values, this term is standardized by dividing by $(n-2)\cdot|\overline{I}|$, where n is the number of total states perceived by the observer.

Furthermore, the model highlights that even interstate behaviors that do not di-

rectly involve the target state can influence how it is perceived. According to Social Identity Theory (SIT), individuals derive part of their identity from the groups to which they belong, so inter-group interactions can influence how individuals perceive both their own group (in-group) and those they are not a part of (out-group). (Tajfel, 1981; Turner & Oakes, 1986). Specifically, individuals observing rising tensions between some members of an in-group and an out-group can intensify hostility towards the entire out-group members, while simultaneously strengthening their emotional attachment to people in their in-group (Hogg, 1992; Gibler et al., 2012). Applying this logic to international politics, individuals may interpret hostile or cooperative behavior between allied and rival states as signals about broader group dynamics. Heightened conflict between members of opposing blocs can reinforce negative images of enemy states and strengthen favorable perceptions of allies, thereby polarizing national images across the ally-enemy continuum. Conversely, when states from opposing groups engage in cooperation, the salience of group boundaries may weaken, prompting individuals to soften preexisting images and contributing to a convergence of national perceptions.

Incorporating this concept, the final component of Equation 1 captures the influence of PTI—that is, behaviors among third-party states not directly involving the target state. In this term, $\hat{B}_{cd,t-1}$ represents the observed interaction between two third-party states c and d. This interaction is weighted by the individual's perception of each state's alignment ($\hat{I}_{a^i c,t-1}$ and $\hat{I}_{a^i d,t-1}$), which together reflect the perceived degree of confrontation between ally and enemy groups in the international system. For example, a hostile interaction ($\hat{B}_{cd,t-1} < 0$) between a perceived enemy ($\hat{I}_{a^i c,t-1} < 0$) and a perceived ally ($\hat{I}_{a^i d,t-1} > 0$) will produce a positive product ($\hat{B}_{cd,t-1}\hat{I}_{a^i c,t-1}\hat{I}_{a^i d,t-1} > 0$), signaling heightened inter-group tension and reinforcing polarized images. Conversely, cooperative behavior between an ally and an enemy ($\hat{B}_{cd,t-1} > 0$) yields a negative product ($\hat{B}_{cd,t-1}\hat{I}_{a^i c,t-1}\hat{I}_{a^i d,t-1} < 0$), which may soften perceived divisions and encourage the reassessment of rigid group boundaries. Because these interactions involve multiple combinations of

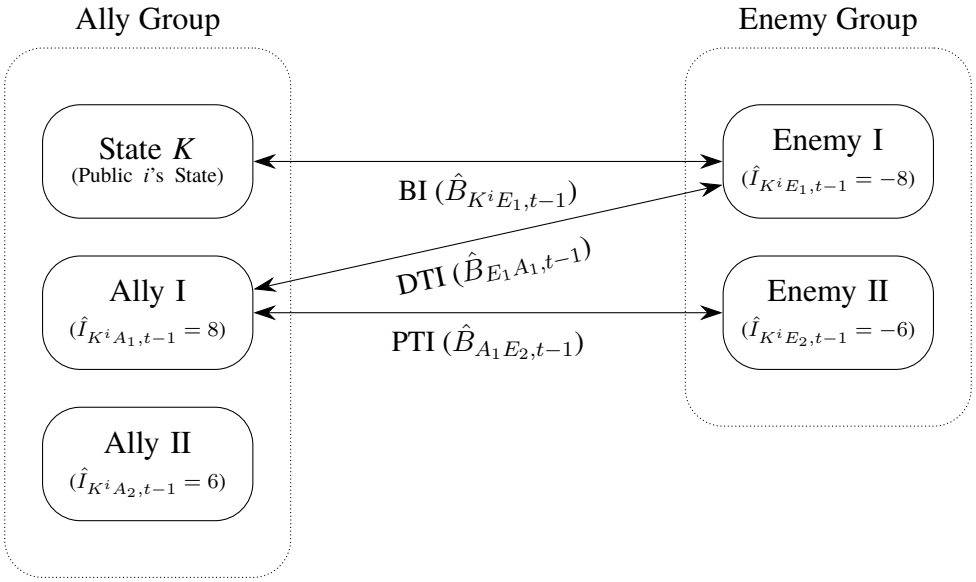
third-party state pairs, the model aggregates them to reflect a broader sense of systemic inter-group dynamics. However, the impact of this aggregated value depends on the observer's prior perception of the target state. If the target is already seen as an enemy ($\hat{I}_{a^ib,t-1} < 0$), heightened inter-group conflict further reinforces its negative image ($\hat{I}_{a^ib,t} < \hat{I}_{a^ib,t-1}$), whereas increased cooperation among third-party states tends to weaken this view ($\hat{I}_{a^ib,t} > \hat{I}_{a^ib,t-1}$). In contrast, if the target is initially perceived as an ally ($\hat{I}_{a^ib,t-1} > 0$), a polarized environment will reinforce its positive image ($\hat{I}_{a^ib,t} > \hat{I}_{a^ib,t-1}$), while a more cooperative international context may reduce its favorability ($\hat{I}_{a^ib,t} < \hat{I}_{a^ib,t-1}$). To account for the number of third-party state pairs and the scaling effect of multiplying three image-related terms, this component is standardized by dividing by $\binom{n-2}{2} \cdot |\overline{I}|^3$.

4. Simulation Analysis and Testable Hypotheses

The image updating model developed in this study is based on the premise that individuals perceive and assign national images to multiple foreign states simultaneously, and that these images are interdependently updated in response to behavioral shocks within the international environment. To address this complexity, this project employs simulation methods that allow for the dynamic modeling of how individuals revise national images across multiple states in response to a given observation of interstate interactions. Specifically, this approach enables the exploration of how different patterns of interstate conflict and cooperation, particularly those involving third-party actors, influence the distribution of perceived allies and enemies. As such, this section introduces the simulation framework—centered on the image updating model—and use its results to develop testable hypotheses regarding the impact of third-party interactions on public perceptions of a foreign adversary.

Figure 2 illustrates the core setup of prior national images and behavioral com-

Figure 2: Simulation Set-up on State Actors and Interstate Interactions

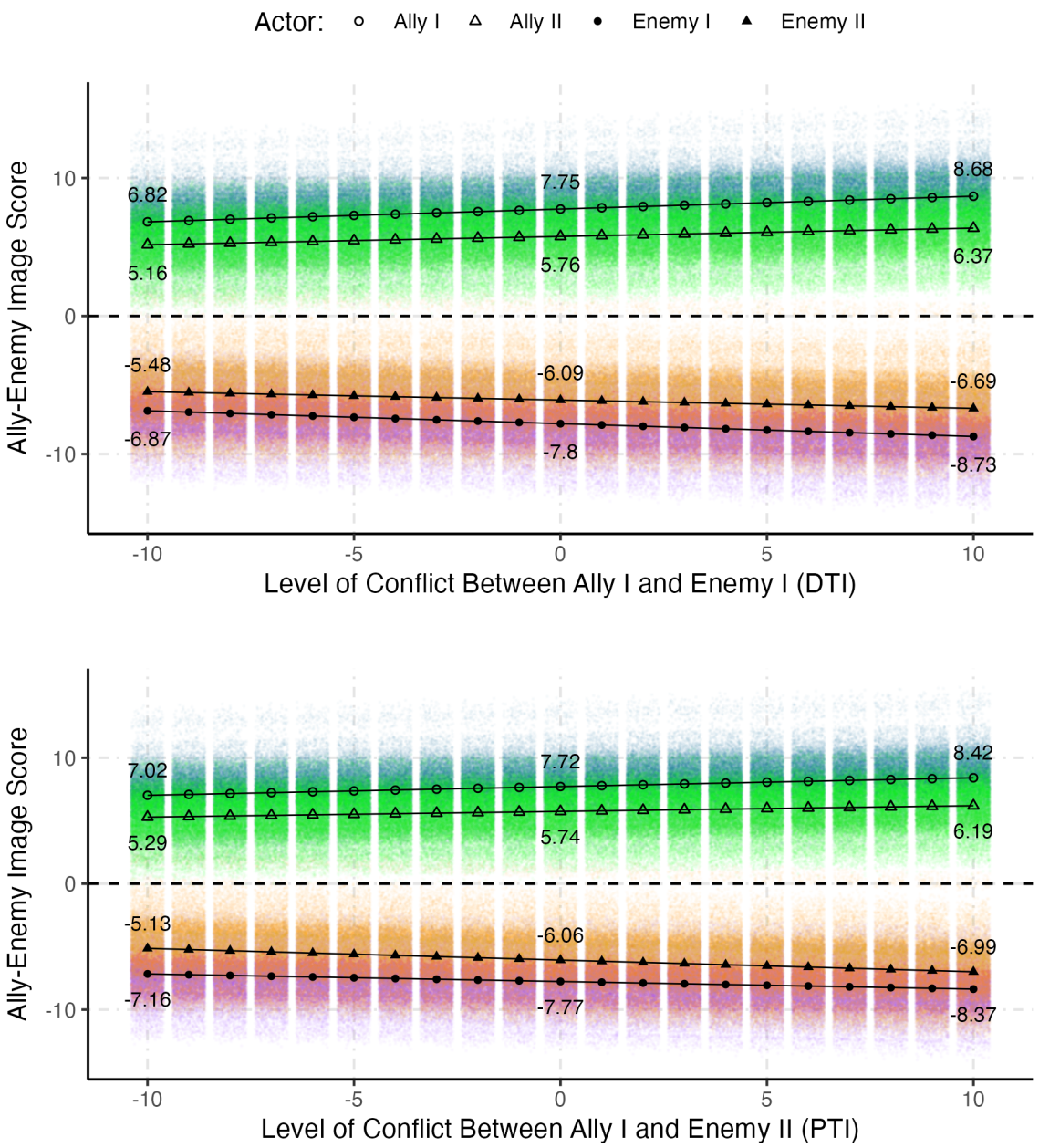


ponents used in the simulation.[3] In this scenario, State K where an individual i is citizen of has a history of international rivalry with State E_1. This adversarial relationship extends beyond the two principal states, shaping perceptions of other international actors. Specifically, individuals classify three additional third-party states as allies or enemies based on their direct involvement in the conflict or their broader political and ideological alignments. As a result, four states play a central role in shaping the perceived international environment: two are seen as enemies—Enemy I ($\hat{I}_{K^iE_1,t-1} = -8$) and Enemy II ($\hat{I}_{K^iE_2,t-1} = -6$)—and two as allies—Ally I (State A_1, $\hat{I}_{K^iA_1,t-1} = 8$) and Ally II ($\hat{I}_{K^iA_2,t-1} = 6$). In addition to these core actors, individuals in State K also form national images of 45 other countries, assessing each along the ally–enemy continuum. Their prior national images are modeled using a generalized logistic probability distribution between -10 and 10 and are randomly drawn for each simulation.

The primary goal of this simulation is to investigate how individuals update their perception of Enemy I ($\hat{I}_{K^iE_1,t}$) in response to two distinct types of third-party interstate interactions: DTI and PTI. The simulation specifically focuses on interactions between states aligned with opposing sides of an international rivalry. So, for DTI,

[3]See Appendix 1 for more comprehensive information on the simulation setup.

Figure 3: Simulation Results on the Effects of DTI and PTI on National Images



Note: The black dots in various shapes, along with the solid black lines, represent the average trend of ally-enemy image scores for the four state actors across the simulations. The minuscule colored dots—blue (Ally I), green (Ally II), purple (Enemy I), and orange (Enemy II)—depict the image scores of each respective actor in individual simulation outcomes.

which involves direct behavior between the target state and a third-party state, the first simulation assesses the effect of varying degrees of cooperation or hostility in interactions between Ally I and Enemy I ($\hat{B}_{E_1A_1,t-1}$). In the case of PTI—which captures interactions between two third-party states that do not directly involve the

target state—the second simulation analyzes the effects of changing levels of cooperation or hostility between Ally I and Enemy II ($\hat{B}_{A_1E_2,t-1}$). To test the effects of these mechanisms, both behavioral variables are systematically varied across integer values from -10 to 10, with higher values representing more cooperative behavior ($\hat{B}_{E_1A_1,t-1}, \hat{B}_{A_1E_2,t-1} \in \mathbb{Z} \cap [-10, 10]$). Meanwhile, all other interactions involving the remaining 48 states—including the bilateral relationship between State K and Enemy I ($\hat{B}_{K^iE_1,t-1}$)—are randomly assigned values drawn from a generalized logistic distribution.

Figure 3 illustrates the outcome of these simulations. The top panel shows how the national images of the four focal states change as a function of varying levels of DTI between Ally I and Enemy I. The results indicate that national images become increasingly polarized as inter-group hostility intensifies. In particular, the enemy image of Enemy I strengthens significantly, with its score dropping from -6.87 under the highest level of cooperation ($\hat{B}_{E_1A_1,t-1} = 10$) to -8.73 under the most hostile condition ($\hat{B}_{E_1A_1,t-1} = -10$). Also, although Enemy II is not directly involved in this interaction, its enemy image also becomes more negative (-5.48 to -6.69). Meanwhile, the ally images of Ally I and Ally II become more positive, increasing from 7.59 to 8.74 and from 5.16 to 6.37, respectively. This growing divergence between the evaluations of allies and enemies reflects the mechanism of image polarization in response to conflict between states aligned with opposing camps. Based on this result, I propose that higher levels of hostility between a rival state and an allied state lead to a stronger enemy image of the rival state (*H1A*). In turn, such heightened enemy image may result in increased domestic resistance to rapprochement efforts. Therefore, I hypothesize that hostility between a rival and an ally leads to greater public opposition to conciliatory policies toward the rival state (*H1B*).

On the other hand, the bottom panel of Figure 3 presents the effects of PTI between Ally I and Enemy II, where the target of interest (Enemy I) is not directly involved. Once again, the results reveal a pattern of polarization: as inter-group

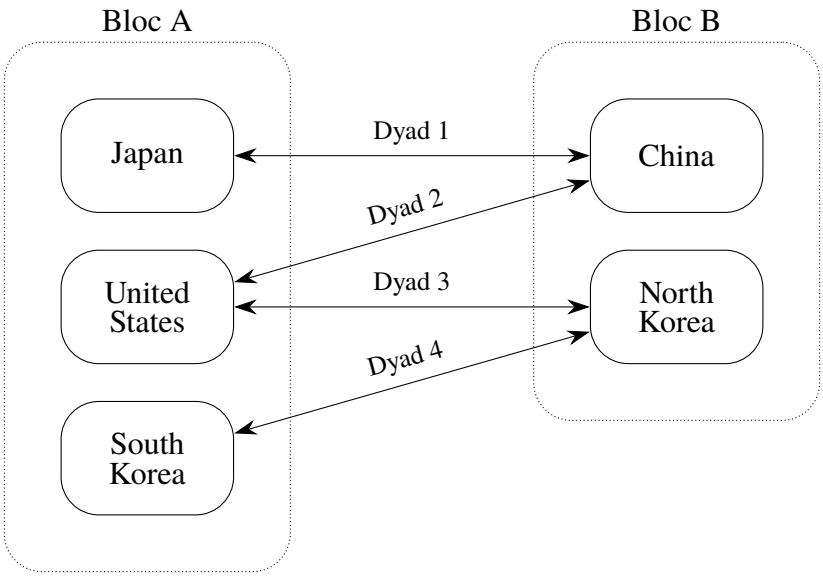
hostility between the third-party states increases, perceptions of allies and enemies diverge more sharply. The ally image of Ally I rises from 7.02 to 8.42, while the enemy image of Enemy II becomes more negative, declining from -5.13 to -6.99. Notably, the enemy image of Enemy I—despite its non-involvement in the interaction—also deteriorates, falling from -7.16 to -8.37. Also, the ally image of Ally II increases from 5.29 to 6.19. This suggests that individuals infer broader patterns of inter-group alignment from third-party conflict, reinforcing preexisting images of allies and enemies. These findings support the hypothesis that greater hostility between two third-party states—one perceived as an ally and the other as an enemy—can intensify negative perceptions of a rival state (*H2A*). Furthermore, such image reinforcement may translate into stronger domestic resistance to engagement with that rival, leading to increased opposition to conciliatory policies (*H2B*).

5. Research Design Overview

To test the theoretical claim, this paper draws on empirical evidence from two original data from a survey experiment (Study 1) and a social media analysis (Study 2). Both studies focus on international rivalries involving East Asian states, taking advantage of the region's dense network of interstate relationships. As illustrated in Figure 4, the strategic landscape in East Asia, centered on five key states—Japan, South Korea, the U.S., China, and North Korea—has been shaped over the past several decades by a division into two opposing blocs. The Democratic Bloc consists of Japan, South Korea, and the U.S., which are closely linked through formal security alliances and shared democratic state systems. The Autocratic Bloc, comprising China and North Korea, stands in opposition to the Democratic Bloc and represents states often positioned against U.S.-led regional influence, both ideologically and strategically. The arrows connecting the groups (*Dyads 1–4*) represent ongoing interstate rivalries: Japan and the U.S. are in adversarial relationships with China (*Dyads 1* and *2*), while the U.S. and South Korea face persistent hostilities

from North Korea (*Dyads 3* and *4*).

Figure 4: Focused International Rivalries in East Asia
And Their Interstate Interactions



Study 1 seeks to generate empirical evidence through the use of an online survey experiment. This methodological approach offers several advantages for the research. First, a survey experiment is well-suited to assess public opinion on a specific foreign policy issue, such as rapprochement, which is typically difficult to directly observe. Moreover, a survey experiment is effective in testing theoretical mechanisms, as it offers the flexibility to incorporate novel measurements—such as national images across multiple states. Last but not least, an experimental approach generally enables a researcher to hold constant characteristics of situations that may confound the result, thus enabling a thorough examination of the underlying causal mechanisms of international interactions involving third-party states.

This experiment aims to evaluate American public attitudes toward pursuing a conciliatory policy toward China. The dependent variable centers on *Dyad 2* in Figure 4, capturing views of the U.S.–China bilateral relationship. In light of the escalating tensions between the U.S. and China, many recent scholars have undertaken surveys to explore how Americans perceive their relationship (Mattes & Weeks, 2019; Myrick & Wang, 2023). However, the US-China rivalry does not exist in isolation on the global stage. The US's involvement in major wars and polit-

Table 1: Research Design Overview

Study	Empirical Method	Public Opinion (Dependent Variable)	Interstate Behavior (Independent Variable)	Hypothesis Test
Study 1	Survey Experiment	**[Dyad 2]** American Public Opinion on China	**[Dyad 1]** Conflict vs. Cooperation Between Japan and China	DTI (*H1A & H1B*)
			[Dyad 4] Conflict vs. Cooperation Between ROK and DPRK	PTI (*H2A & H2B*)
Study 2	Text Analysis of Social Media	**[Dyad 4]** S.Korean Public Opinion on DPRK	**[Dyad 3]** Conflict vs. Cooperation Between the US and DPRK	DTI (*H1A & H1B*)
		[Dyad 1] Japanese Public Opinion on China		PTI (*H2A & H2B*)

ical disputes in the East Asia region, especially through the Korean War and Cold War, Americans are assumed to perceive Japan and South Korea as allies, while viewing China as a primary adversary and North Korea as China's strategic partner.(Friedhoff, 2020; Kafura, 2022).[4] Within this structure, two types of inter-group interactions emerge beyond the direct U.S.–China dynamic. The first illustrated by *Dyad 1*, involves a direct third-party interaction (DTI) between Japan and China. The second, shown in *Dyad 4*, reflects a peripheral third-party interaction (PTI) between North and South Korea. The study aims to experimentally investigate how exposure to these third-party behaviors influences American perceptions of China.

Although experimental designs are useful for uncovering causal mechanisms, questions regarding their external validity persist. In particular, while experiments enable controlled manipulation of treatments and strong causal inference, they are less effective at capturing how individuals respond in real-world settings (Barabas & Jerit, 2010). To address this, Study 2 provides observational evidence using social media data retrieved from South Korea and Japan. Though social media may

[4]According to the survey of the Chicago Council on Global Affairs in October 2022 (Kafura, 2022), American ratings of South Korea and Japan on a 0-100 feeling thermometer scale averaged 72 and 62, respectively, while those of China and North Korea are 32 and 20. Also, over the past decade, the perspectives of Americans toward these two groups of countries have become increasingly divergent (Friedhoff, 2020).

lack representativeness within each country, it offers access to natural, unsolicited expressions of public sentiment at scale.

Unlike Study 1, which manipulated the behavior of third-party dyads, Study 2 centers on a single behavioral dyad while varying the target dyads used to measure public opinion. The independent variable in this study is the significant shift in U.S.–North Korea relations between 2017 and 2018, corresponding to *Dyad 3* in Figure 4. The analysis then explores how this behavioral change influences public attitudes in South Korea and Japan toward their respective rival states, North Korea and China. In the South Korean case, represented by *Dyad 4*, North Korea is traditionally perceived as the primary threat, while the U.S. is regarded as a long-standing ally, Accordingly, the behavioral shift in U.S.–North Korea relations serves as a DTI treatment, which may influence South Korean attitudes toward North Korea. In contrast, *Dyad 1* captures the Japanese context, where China is viewed as the principal rival. Given Japan's alliance with the U.S. and China's close association with North Korea, the changing U.S.–North Korea dynamic functions as a PTI shock—potentially shaping Japanese views of China and affecting support for a conciliatory foreign policy stance.

6. Study 1: Survey Experiment on US-China Rivalry

In August 2023, I fielded a pre-registered survey experiment to a nationally representative sample of 1,412 American adults through Lucid Theorem.[5][6] The survey experiment proceeded in the following sequence. At the initial stage, respondents were asked a series of demographic questions, including their gender, age, religion, household income, education, partisan identity, and political ideology. Next, participants were assigned randomly to either a control condition or one of the four

[5]The experimental design in this paper was approved by Institutional Review Board and was pre-registered with Open Science Forum (OSF) before data collection.

[6]The main analysis of my study focuses on 1,241 responses from participants who successfully passed the survey's pre-treatment attention check. See more details in Appendix

treatment conditions. As Table 1 illustrates, the treatment groups were categorized based on two dimensions: (1) whether observed behaviors are characterized as hostile or cooperative, and (2) whether these interactions involve a DTI between China and Japan or a PTI between South Korea and North Korea.

Table 2: Treatment Conditions and Video Clip Sources

	Conflict	**Cooperation**
DTI (Japan-China)	**Treatment 1 (N = 247)**: "Diaoyu-Senkaku islands spat deepens as Japan warns China over coastguard ships in East China Sea" (published by SCMP on 2020/10/14)	**Treatment 2 (N = 280)**: "Japan's foreign minister to visit China after leaders Fumio Kishida and Xi Jinping meet at APEC" (published by SCMP on 2022/11/18)
PTI (Sotuh-North Korea)	**Treatment 3 (N = 249)**: "South Korea opens fire as North Korea drones cross the border" (published by SCMP on 2022/12/27)	**Treatment 4 (N = 274)**: "North and South Korea restore communication and military hotline after 2 months of silence" (published by SCMP on 2021/10/05)

Subjects in these treatment groups were exposed to a video clip that showcases an international affair that highlights interstate behaviors relevant to each category.[7] Specifically, the first two treatments focus on interactions between Japan and China, with Treatment 1 depicting a hostile maritime dispute over the Senkaku/Diaoyu Islands in 2020, and Treatment 2 highlighting diplomatic engagement following high-level bilateral talks in 2022. The third and fourth treatments shift focus to South Korea and North Korea. Treatment 3 presents a hostile incident involving North Korean drones violating South Korean airspace in 2022, prompting a military response, while Treatment 4 features a cooperative moment in which North and South Korea reestablish communication through a restored military hotline in 2021. I used real media content because it is believed to enhance participants' engagement and comprehension, resulting in more accurate and meaningful responses that closely mirror real-life reactions. In this regard, the video clips were selected from news coverage published by the South China Morning Post (SCMP) on its publicly available YouTube channel. The decision to use the SCMP, instead of other US-based broadcasting companies, was made to (1) unify the source of informa-

[7] For more details about each video clip, refer to Appendix 2.2

tion for all four treatment groups and (2) minimize any potential biases caused by the participants' pre-existing attitudes towards the media source. The participants allocated to the control group did not view any video clip (N = 191).

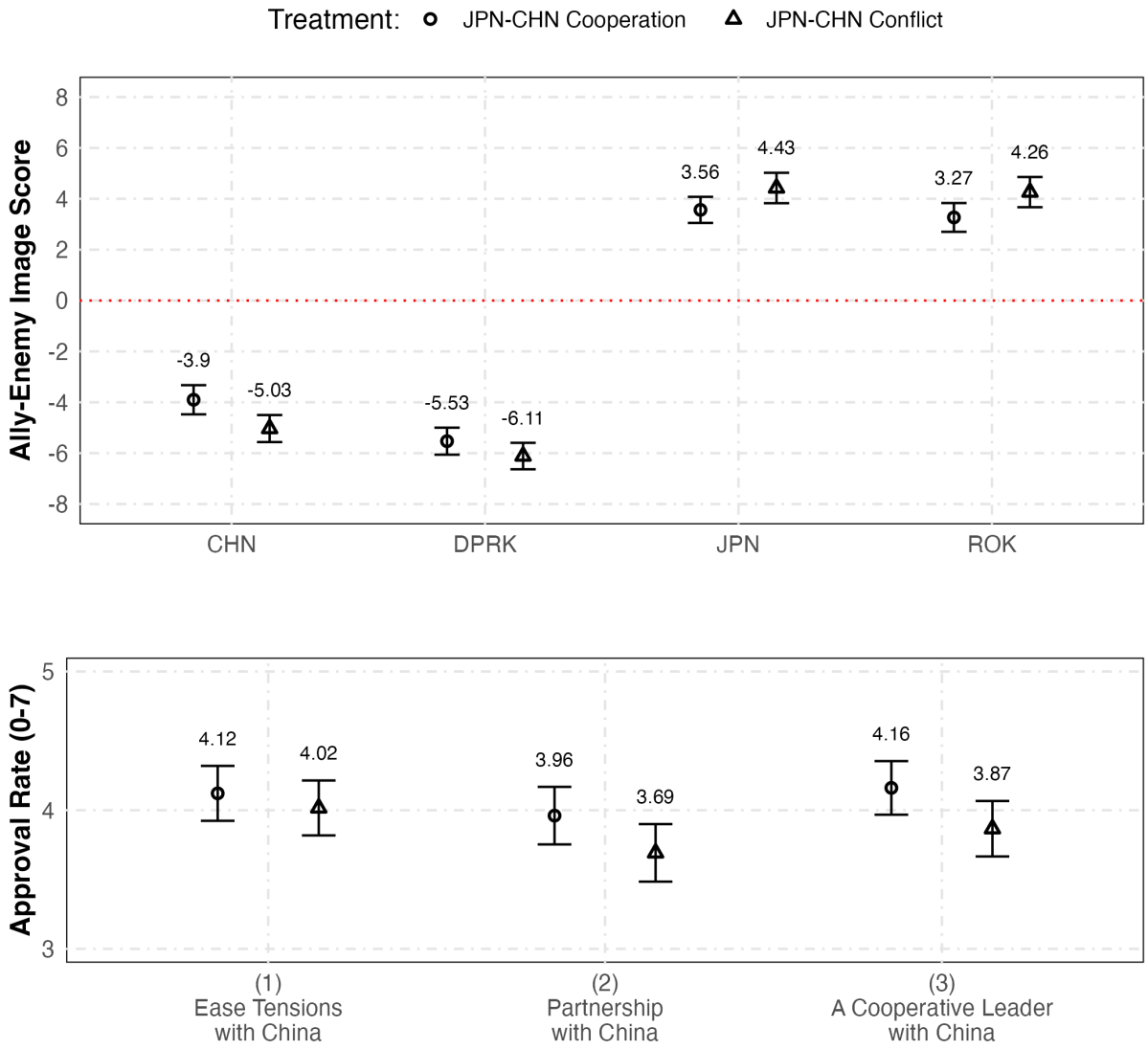
After the treatment stage, two types of dependent variables were measured. First, the survey asked respondents about to what extent they perceive four countries–China, Japan, South Korea, and North Korea–as allies or enemies.[8] To directly measure international images, the following two questions were asked: (1) "On a scale of 0 to 10, to what extent you perceive each of the listed countries as an enemy to the US?" and (2) "On a scale of 0 to 10, to what extent you perceive each of the listed countries as an ally to the US?" When these questions were given, the names of the four countries were displayed in a random order with a bar-type grader. By subtracting the enemy image score from the ally image score, an ally-enemy score is constructed, aligning with the scale employed in the simulation model (ranging from -10 to 10). Second, the survey measured approval of a cooperative policy toward China by asking participants how much they agree with three statements, using a 7- point scale ranging from "strongly disagree" to "strongly agree." These three statements were: (1) The U.S. Government needs to ease tensions in its relationship with China, (2) "Enhancing the partnership with China is likely to be beneficial to the national interests of the U.S.," and (3) "I endorse a leader who promotes cooperation with China rather than engaging in conflicts with it." After completing the survey, participants were debriefed by being provided with the original sources of the video clips shown during the treatment phase.

The initial analysis of the survey experiment examines how American respondents perceive four key countries—Japan, South Korea, China, and North Korea—using an ally-enemy image scale. The findings align with previous survey results, revealing a clear distinction in perceptions between U.S. allies and adversaries in East

[8] Additionally, drawing on prior definitions of ally and enemy images (Herrmann et al., 1997; Alexander et al., 2005), two supplementary questions—one assessing threat perception and the other goal compatibility—were included to indirectly measure enemy image. Further details are provided in Appendix 2.

Asia. Japan (3.54) and South Korea (3.26) receive favorable ratings, while China (-4.28) and North Korea (-5.58) are viewed negatively. These patterns support the assumption that Americans clearly differentiate between allied and rival nations when forming opinions about international relations in the region.

Figure 5: Treatment Effects of Interactions between China and Japan



Note: All error bars in the figures presented in this paper represent 95% confidence intervals.

The main focus of this study is to compare responses between the treatment conditions of conflict and cooperation involving third-party states.[9] First, the anal-

[9]While comparing treatment and control group responses is important, such analysis may be misleading due to the "us-versus-them" framing present in all treatment conditions. Even cooperative treatments may reinforce competitive dynamics by presenting two states in a rivalry. See Appendix 2 for additional analyses.

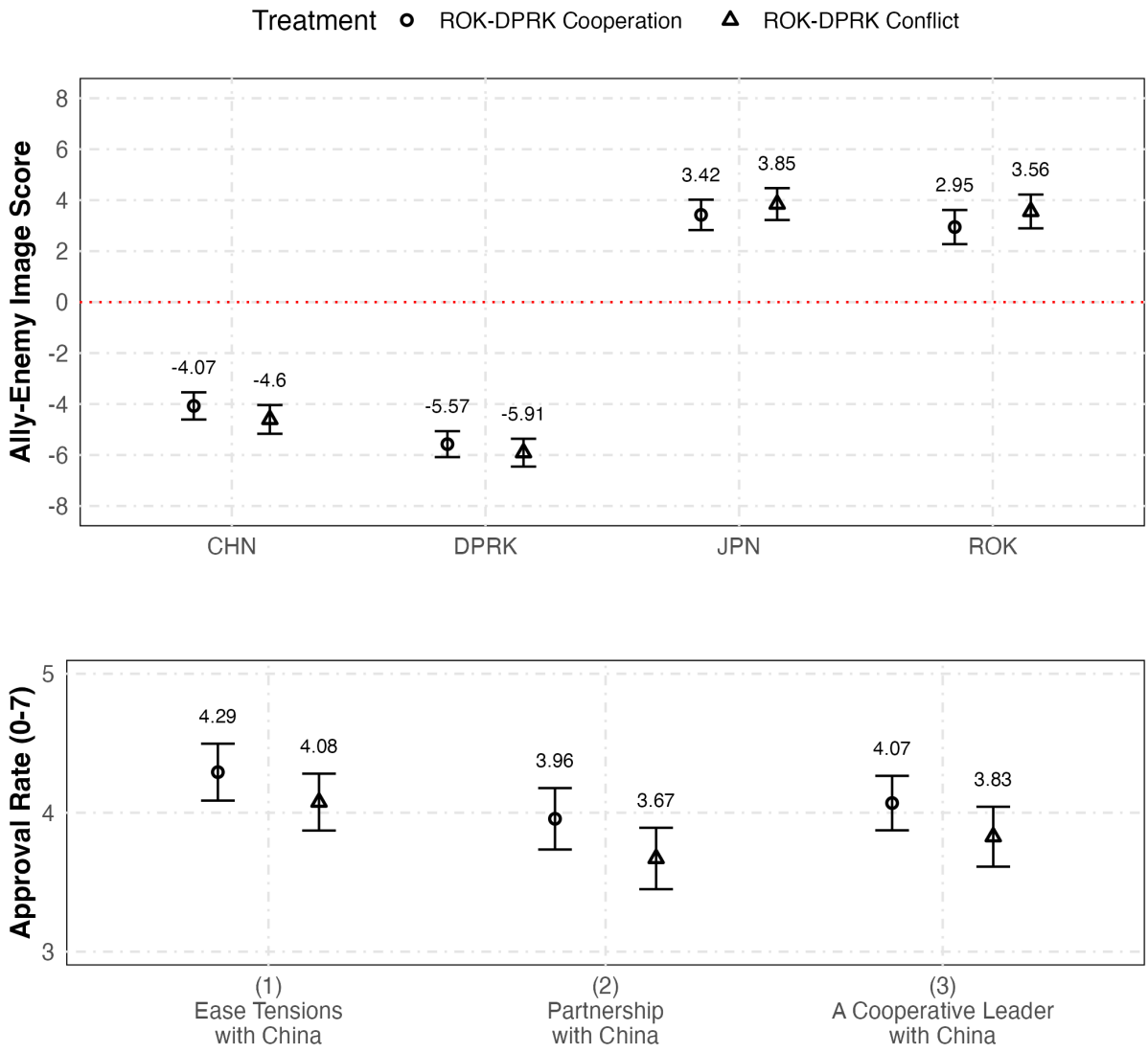
ysis begins by examining the effects of interstate behaviors between an ally state and a rival state (DTI), respectively represented by Japan and China in the experiment. The experiment provides compelling evidence in support of the hypothesis that their enhanced conflict can mitigate an enemy image that Americans perceive from China H1A. Looking at the top panel in Figure 5, participants who viewed the video depicting tensions between China and Japan over the Senkaku/Diaoyu Islands were more likely to characterize China as an enemy, compared to those who watched footage of the APEC meeting between President Xi Jinping and Prime Minister Fumio Kishida. Specifically, China's enemy-ally image score decreases by 1.13 points ($p<0.01$), representing a 22.47% rise in its perception as an enemy in contrast to being perceived as an ally.

According to the theory, this perceptional shift occurs in tandem with the polarization of ally and enemy images that arises from observing inter-group conflict. When individuals witness hostility between perceived allies and enemies, their national images become more extreme—strengthening negative views of enemies and positive views of allies. If this mechanism is valid, we should observe not only a shift in the rival's image but also reinforcement of other ally and enemy images. As shown in Figure 5, the survey respondents who viewed the Senkaku/Diaoyu conflict rated North Korea more negatively by 0.58 points ($p = 0.12$), while their perceptions of Japan and South Korea as allies declined by 0.86 ($p = 0.03$) and 0.99 ($p = 0.02$) points, respectively. This pattern supports the notion that inter-group conflict sharpens group boundaries, contributing to broader shifts in perceptions of the international landscape.

Moreover, the results demonstrate significant evidence that supports H1B, which claims for the negative impact of a hostile DTI on public support for a conciliatory policy. This pattern is reflected in the second panel of Figure 5. Respondents exposed to the conflict between China and Japan expressed lower approval for partnering with China, showing a 0.27-point drop compared to those in the cooperation condition ($p = 0.07$), equivalent to an 6.82% decrease in approval. Furthermore,

the conflict group exhibited a 0.29-point increase in disapproval toward U.S. leaders who promote peaceful engagement with China relative to the cooperation group ($p = 0.04$), indicating a 6.97% difference. While the conflict treatment group also showed a slightly higher likelihood of supporting a de-escalatory foreign policy compared to the cooperation group, this difference was not statistically significant ($p = 0.47$).

Figure 6: Treatment Effects of Interactions between North and South Korea



My next analysis centers on the impact of international interactions that only involve two third-party states (PTI). As shown in Figure 6, this assessment is based on a comparison between the third and fourth treatment groups, which were re-

spectively exposed to video footage depicting hostile and cooperative exchanges between North and South Korea. This comparison reveals a pattern that mirrors the earlier findings, albeit to a lesser extent, offering suggestive support for H2A. The results indicate that participants who viewed the hostile encounter between North and South Korea reported a decrease of 0.53 points in their ally-enemy image score on China—representing a 13% more negative perception—relative to those who observed the cooperative interaction. However, this difference does not reach statistical significance at the 0.1 threshold (p = 0.182).

Also, the findings offer suggestive evidence that hostile inter-group interactions drive the divergence of national images. Echoing the earlier case involving Japan and China, participants who witnessed conflict between North and South Korea evaluated the listed states in more polarized terms than those who observed cooperative engagement between the two. Specifically, compared to the cooperation condition, exposure to inter-Korean conflict increased perceptions of North Korea as an enemy by 0.33 points (p = 0.38), and heightened the identification Japan and South Korea as allies by 0.42 (p = 0.37) and 0.61 points (p = 0.2), respectively. These shifts suggest that witnessing conflict between opposing bloc members may intensify perceived inter-group competition, prompting a broader reassessment of national images.

Finally, the survey provided supportive evidence for H2B, which addresses the impact of PTIs on public attitudes toward a conciliatory policy—though the evidence is less robust than that for DTIs. As shown in the first graph of the last row in Figure 6, respondents who viewed a video depicting hostile relations between North and South Korea expressed lower support for pursuing a partnership with China, with a 0.29-point decrease compared to those exposed to inter-Korean cooperation (p = 0.06), representing an 7.32% reduction. Additionally, the conflict group showed a 0.22-point (5.13%) decline in approval for easing tensions with China and a 0.24-point (5.9%) drop in support for leaders favoring a cooperative approach—both effects failing to reach statistical significance at the 0.1 level (p =

0.14 and 0.1, respectively).

7. Study 2: Social Media Analysis in South Korea and Japan

This study examines how U.S.–North Korea interactions between 2017 and 2018 shaped public perceptions in South Korea and Japan, particularly regarding their respective international rivals, North Korea and China. During this two-year period, the U.S.–DPRK relationship underwent dramatic shifts, characterized first by escalating tensions and later by unprecedented diplomatic engagement. In 2017, North Korea conducted multiple missile tests, including long-range launches capable of reaching the U.S.. In response, then-President Donald Trump issued a stark warning, declaring that North Korea would be met with "fire and fury like the world has never seen." The situation intensified further when North Korea conducted its sixth nuclear test, prompting the Trump administration to adopt a "maximum pressure" strategy that combined strict sanctions with a strong emphasis on military deterrence. However, 2018 marked a notable change in tone. The year began with North Korea's participation in the Winter Olympics in South Korea, signaling a temporary thaw in tensions. This was followed by a series of diplomatic efforts, including visits by U.S. Secretary of State Mike Pompeo to Pyongyang and culminating in the historic Singapore summit between Kim Jong-un and Donald Trump. This meeting—the first between a sitting U.S. president and a North Korean leader—resulted in a joint statement in which North Korea pledged to work toward the complete denuclearization of the Korean Peninsula.

To assess public attitudes toward a rival state, Twitter (now known as X) is used as the primary platform for collecting social media data. The initial step involved collecting tweets that might reflect public sentiment in South Korea and Japan regarding their respective foreign adversaries, North Korea and China.[10] This process

[10]For the Korean case, the search focused on Korean-language posts that referenced terms such as "North Korea" (북한 or 대북) or "Kim Jong-un" (김정은), posted between 2017 and 2018. Similarly, for the Japanese case, tweets were collected that mentioned "China" (中国, 支那, or 對中) or "Xi Jinping" (習近平).

yielded a total of 729,867 tweets and retweets from South Korea and 366,543 from Japan. Next, this study employed a supervised machine learning model to construct the dependent variables measuring negative perceptions toward rival states in South Korea and Japan. To train the model, 10,000 Korean and 5,000 Japanese tweets were randomly sampled and manually annotated based on attitudes toward North Korea and China, respectively.[11] Drawing on the conceptualization of enemy images as rooted in perceived threats and goal incompatibility, tweets were assigned a value of 1 if they described the adversary as a security threat or expressed conflicting national goals or interests. A binary variable, labeled "enemy image," was then created and coded as 1 if at least one of these two dimensions was present. Additionally, a second dependent variable, "anti-dovish sentiment," was developed to capture criticism of domestic political actors for adopting conciliatory stances toward the foreign adversary. Tweets expressing such criticism were also coded as 1. The full coding scheme used in this annotation process is detailed in Appendix 3.

Using the annotated data, a predictive model was trained to estimate the probability that each tweet expresses a particular attitude based on its textual content. For Korean-language tweets, the normalization and tokenization processes were carried out using KoNLPy (Park & Cho, 2014), a widely used natural language processing toolkit for Korean text analysis. A similar preprocessing procedure was applied to Japanese tweets using Nagisa (Matsuzaki, 2020), a Japanese text segmentation tool optimized for Python. To perform the classification task, two supervised machine learning algorithms—Random Forest and Support Vector Classification (SVC)—were employed, both of which are commonly used for text classification problems.[12]

To construct an independent variable, I employed the Integrated Conflict Early Warning System (ICEWS). Initially developed to assist U.S. policy analysts in fore-

[11]The annotation of Korean tweets was conducted by the author, a native Korean speaker. For the Japanese corpus, two native Japanese speakers assisted in the annotation process.

[12]For further details on the annotation procedure and model performance, see Appendix 3.1 and 3.2.

casting international developments (Ward et al., 2013; Boschee et al., 2018), the ICEWS dataset has since gained traction among international relations scholars as a valuable source of empirical data (Minhas et al., 2016; Myrick & Wang, 2023). This dataset is particularly well-suited for this study for two key reasons. First, it captures the intensity of hostile and cooperative actions—scored from -10 (most hostile) to +10 (most cooperative)—carried out by 167 governments against both state and non-state actors, which enables the systematic measurement of U.S.–North Korea interactions during 2017 and 2018. Second, because ICEWS compiles events from news media sources, it is likely that the foreign policy behaviors documented—particularly those involving the U.S. and North Korea—were visible to and recognized by the South Korean public.

Figure 7: US-DPRK Interactions between 2017 and 2018 (ICEWS)

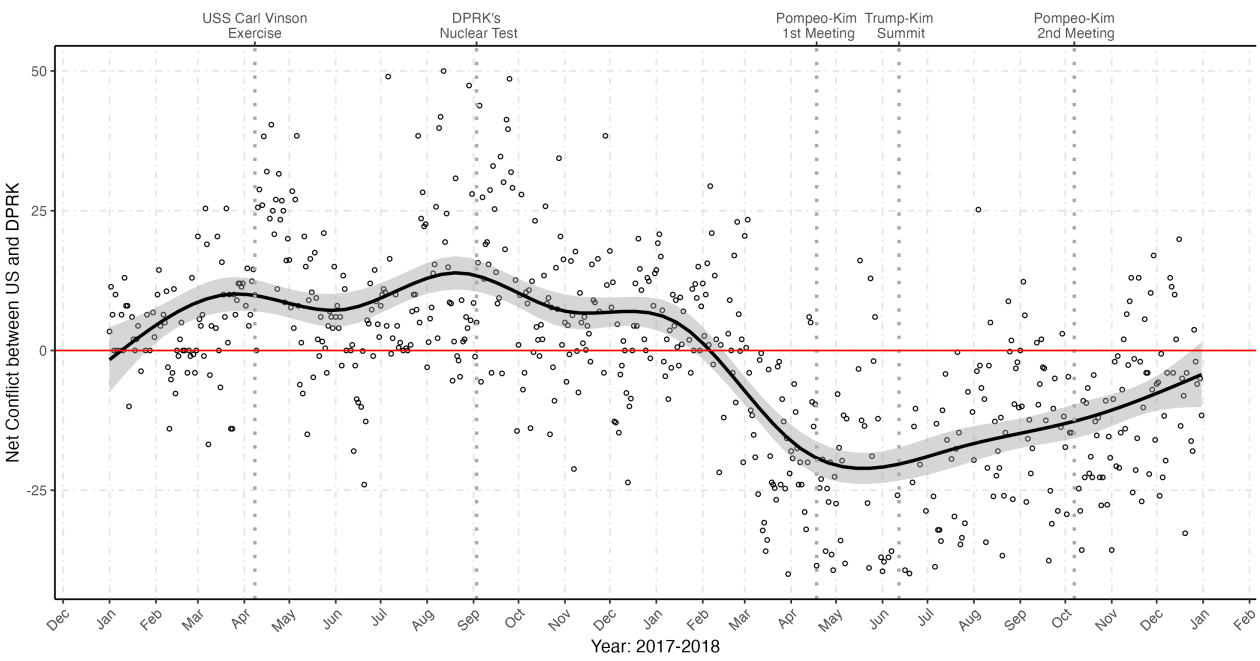


Figure 7 displays the events that are documented in the ICEWS, representing the level of hostile and cooperative interactions between the U.S. and North Korea during 2017 and 2018. The vertical axis of this graph indicates the level of net conflict between the U.S. and North Korea, which is calculated by subtracting the intensity of cooperative actions from that of hostile actions. Each empty dot indi-

cates the accumulated level of net cooperation between two countries on each day during the two-year period. The smooth line in this figure is a Generalized Additive Model (GAM) line that shows the average non-linear trend of net cooperation between the two countries with 95% confidence interval. The graph confirms the dramatic shift in the US-DPRK relationship: it was characterized by a period of heightened tensions in 2017 and a series of diplomatic engagements in the following year, including historic summits in Singapore in June 2018.

To examine the impact of international events on public sentiment, I construct the main explanatory variable using two temporal windows. The first captures the net intensity of conflict on the day prior to each tweet's publication, while the second reflects the cumulative net intensity over the week leading up to the tweet. For the latter, for example, if a tweet referencing North Korea or Kim Jong-un was posted on January 8th, 2018, the weekly variable measures the level of net cooperation between the U.S. and North Korea from January 1st to 7th, 2018, and examines its association with whether the tweet portrays North Korea as an enemy or criticizes the Moon administration's pro-DPRK policy. This dual approach allows me to assess both immediate and sustained effects of international behavior. Similarly, I include control variables capturing the net conflict between (1) North and South Korea and (2) China and Japan over the same period, as these regional dynamics may also shape how South Korean and Japanese publics perceive North Korea and China, respectively.

The first part of the analysis presents evidence from South Korea, which gives substance to the first set of hypotheses on the impact of interstate behaviors involving an ally and a rival state (PTI). In general, the findings are aligned with both H1A and H1B, showing that an increasing level of hostile behaviors between the U.S. and North Korea increases the South Korean public's restraints against reconciliation towards North Korea. The top-left panel of Figure 8 illustrates the relationship between U.S.–DPRK conflict and the likelihood that South Korean tweets depict North Korea as an enemy. The steep upward slope of the regression line, along with

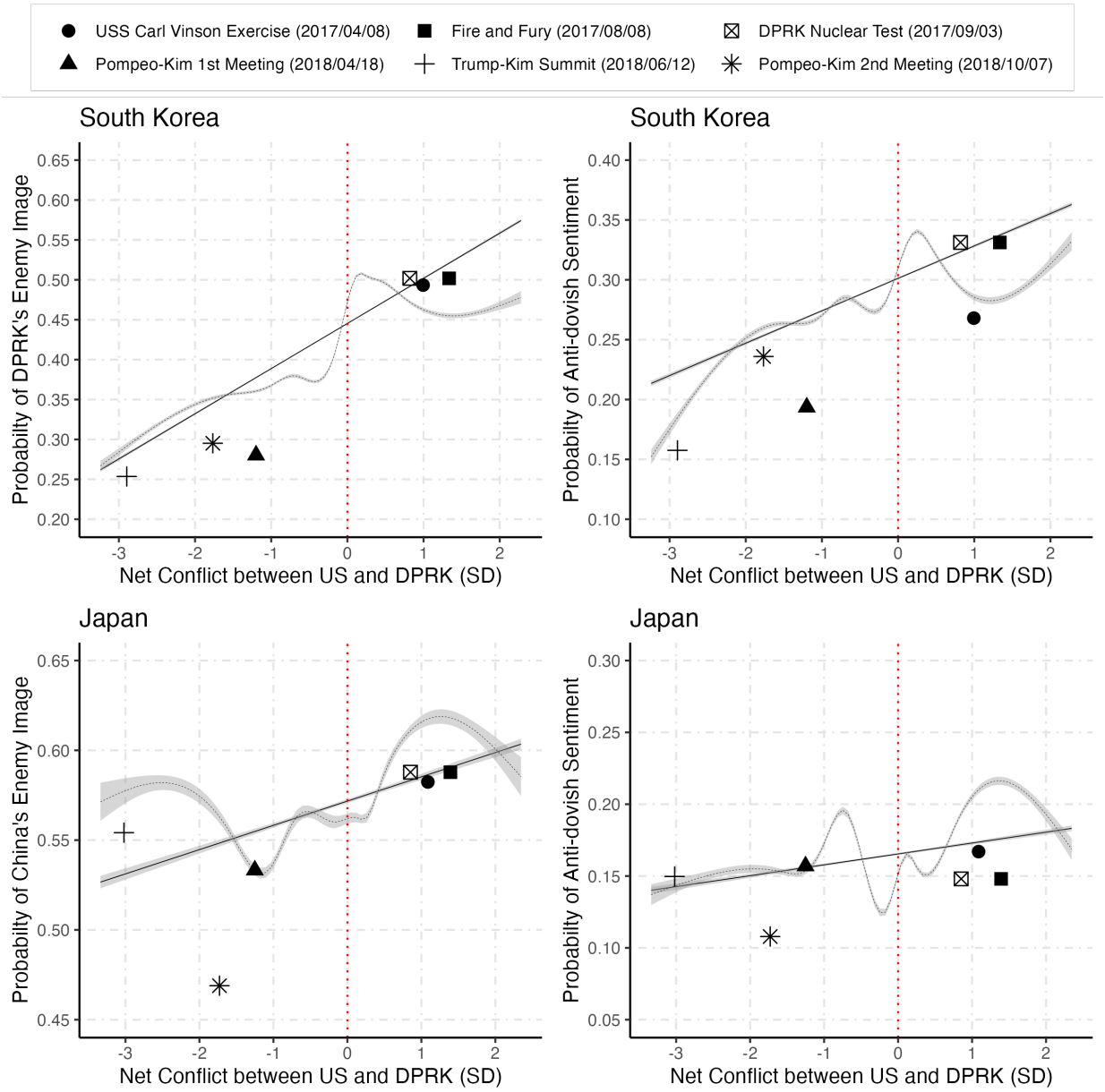
its narrow confidence interval, indicates a statistically significant positive association between the two variables. Based on the bivariate model using a one-week window, a one-standard-deviation increase in conflict intensity between the U.S. and North Korea is associated with a 5.66% increase ($p < 0.01$) in the probability that South Korean tweets characterize North Korea as an enemy.[13] This pattern is also evident in real-world events: in August 2017, following President Trump's "fire and fury" statement, 50.19% of relevant South Korean tweets portrayed North Korea negatively, compared to just 25.36% during the period of U.S.–DPRK rapprochement, marked by Trump's summit with Kim Jong-un in Singapore in June 2018.

Additionally, the top-right panel of Figure 8 demonstrates that hostile interactions between the U.S. and North Korea also intensified domestic criticism of pro-DPRK political groups in South Korea. The graph indicates that a one-standard-deviation increase in U.S.–DPRK conflict corresponds to a 2.71% increase in such domestic backlash ($p < 0.01$). Consistent with the previous finding, anti-dovish sentiment peaked around the time of President Trump's "fire and fury" statement in 2017, with 33.12% of relevant tweets expressing criticism of pro-DPRK stances—17.37 percentage points higher than during the period surrounding the Trump–Kim summit in 2018, when such criticism fell to 15.76%.

My analysis next focuses on the case of Japan, where I evaluate empirical support for H2A and H2B. Overall, the findings from Japanese Twitter data lend support to the hypotheses suggesting that conflict between an ally and a rival's ally (PTI) can heighten domestic opposition to rapprochement. The bottom-left panel of Figure 8 illustrates how U.S.–North Korea cooperation influences the likelihood that Japanese tweets portray China as a threat. Mirroring the trend observed in South Korea, the upward slope of the regression line indicates that higher levels of U.S.–DPRK conflict are associated with an increased tendency among Japanese

[13] These results are robust to alternative model specifications using different time windows and additional control variables accounting for other interstate interactions. For further details, see Appendix 3.3.

Figure 8: The Effects of US-DPRK Interactions on Public Sentiments in S. Korea and Japan



Note: The solid lines in this graph represent the predicted line from the OLS model, whereas the dashed curved lines correspond to the GAM smoothed line. The ribbons around the regression lines indicate the 95% confidence interval. The data points associated with the listed events represent the average probability within ten days after the occurrence of each respective event.

users to describe China as an enemy. However, the effect is more modest than in the South Korean case: a one-standard-deviation rise in U.S.–North Korea conflict corresponds to a 1.35% increase in negative portrayals of China. Finally, consistent with theoretical expectations, hostile interactions between an ally and a rival's ally appear to strengthen domestic resistance to rapprochement. As shown in the

bottom-left panel, a one-standard-deviation increase in U.S.–DPRK conflict is associated with a 0.76% increase in the likelihood of tweets criticizing pro-China political actors in Japan.

8. Conclusion

Returning to the comparison between Presidents Roh Moo-hyun and Moon Jae-in introduced earlier, this paper offers a key insight into the divergent public responses to their respective peace initiatives. Although both leaders pursued similar engagement strategies with North Korea, South Koreans interpreted these efforts differently based on the broader international context. During Roh's presidency, public skepticism was shaped in part by the post-9/11 global security climate, particularly the U.S. framing of North Korea as part of the "Axis of Evil" (Cha, 2002), which heightened perceived threats and made Roh's conciliatory stance appear misaligned with prevailing global narratives. In contrast, Moon's initiatives benefited from a supportive international environment, marked by the Trump administration's summit diplomacy with Kim Jong-un (Brookings Institution, 2018) and endorsements from the European Union (Menegazzi, 2018). These favorable signals reduced the perceived threat of North Korea and enhanced the credibility of Moon's outreach, contributing to broader public support for engagement.

Other historical cases, though varied in context and scope, reveal a shared pattern. IIn post-war Europe, long-standing hostility between France and Germany gradually diminished through the cooperative efforts of neighboring states—such as Italy, Belgium, the Netherlands, and Luxembourg—within the framework of the European Coal and Steel Community. The U.S.–China rapprochement in the early 1970s similarly contributed to the normalization of China–Japan relations, culminating in the Shanghai Communiqué. The end of the Cold War further marked a turning point, ushering in a more fluid and multipolar international order that enabled new openings for diplomatic normalization. This shift facilitated rapproche-

ment in several cases, including between South Korea and Russia, Turkey and Greece, and India and China, as reduced systemic tensions allowed former adversaries to reassess their relationships. Additionally, regional initiatives and third-party signaling—such as those seen during the 2004 SAARC summit—helped ease India–Pakistan hostilities and fostered opportunities for engagement. While direct evidence of public opinion driving these diplomatic shifts is limited, applying an analytical lens beyond a rivalry dyad remains valuable for understanding how broader international dynamics create the conditions under which peace initiatives can gain or lose domestic traction.

In addition to its historical explanation, this study provides several key policy and practical insights for fostering public support for peace between international rivals. The findings demonstrate that such support is shaped not only by direct bilateral interactions but also by how individuals interpret the behavior of third-party states. Policymakers engaged in peacebuilding efforts, therefore, must look beyond the immediate dyad and consider how allied and adversarial third-party actions influence public perceptions of conflict and cooperation. For instance, when an ally engages in hostile behavior toward a rival, it may unintentionally harden domestic public attitudes—even if the rival signals a willingness to reconcile. These dynamics suggest that sustaining public backing for rapprochement requires more than effective bilateral diplomacy; it also demands coordinated alliance management and broader messaging strategies that take into account the ripple effects of third-party interstate behavior.

Moreover, the current political climate may present a timely opening for engagement. In the wake of the Trump administration's "America First" posture and the erosion of traditional ideological cleavages in foreign policy, citizens may be less anchored to long-standing partisan narratives and more attentive to dynamic cues from the international environment. As conventional worldviews lose salience, individuals may become more open to reassessing rival states based on observable interstate behaviors rather than inherited ideological positions. This environment

creates a strategic window for reframing adversaries and fostering public support for peace initiatives. In such a context, visible and consistent cooperation across group boundaries could play a critical role in reshaping national images and generating the domestic political space needed to pursue rapprochement.

While this paper offers a novel theoretical account of how a competitive international environment shapes public opinion on rapprochement, several important directions for future research emerge from its scope conditions. One such condition is the assumption that individuals, particularly in the context of international rivalry, tend to rely on simplified binary categorizations of foreign states—as either allies or enemies. While this framework provides analytic clarity and supports the modeling of image-based opinion updating, it may understate the complexity of real-world public perceptions. In practice, individuals may hold more ambivalent or multidimensional views of foreign states, distinguishing, for example, between economic cooperation and military threat, or between a regime and its people. Future work could address this limitation by developing more nuanced measures of national image that account for issue-specific or conditional perceptions, and by exploring alternative categorizations beyond the ally–enemy dichotomy. Such efforts would help capture the layered nature of public opinion and enhance our understanding of how individuals interpret signals in a complex and evolving international system.

In addition, this study focuses primarily on inter-group interactions—those between states aligned with opposing blocs—as the source of image polarization. Yet, interactions within alliance networks may also shape public perceptions in meaningful ways. Strong cooperation and cohesion among allied states can reinforce group identity, potentially heightening animosity toward adversaries. Conversely, tensions or fractures within an alliance may erode this sense of unity, thereby softening enemy images and dampening support for rivalry. Future research would benefit from investigating the dynamics of intra-group behavior and their effects on the stability of ally–enemy distinctions. Doing so would extend the theoreti-

cal framework proposed here and shed further light on how alliance politics and internal divisions influence public attitudes in international relations.

References

Albuyeh, R., & Paradis, M. (2018). Thawing rivalries and fading friendships: an experimental approach to rapprochement and alienation. *Political Psychology*, *39*(4), 811–827.

Aldrich, J. H., Gelpi, C., Feaver, P., Reifler, J., & Sharp, K. T. (2006). Foreign policy and the electoral connection. *Annu. Rev. Polit. Sci.*, *9*, 477–502.

Alexander, M. G., Levin, S., & Henry, P. J. (2005). Image theory, social identity, and social dominance: Structural characteristics and individual motives underlying international images. *Political Psychology*, *26*(1), 27–45.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review*, *104*(2), 226–242.

Bar-Tal, D., & Halperin, E. (2013). The psychology of intractable conflicts. In *The oxford handbook of political psychology.* Oxford University Press.

BBC News. (2017). *Moon jae-in: South korea's new president, in his own words.* Retrieved from https://www.bbc.com/news/world-asia-39860158 (Accessed: 2025-04-15)

Beardsley, K. (2024). The coevolution of networks of interstate support, interstate threat, and civil war. *The Journal of Politics*, *86*(4), 1418–1430.

Berinsky, A. J. (2009). *In time of war: Understanding american public opinion from world war ii to iraq*. University of Chicago Press.

Bilali, R. (2010). Assessing the internal validity of image theory in the context of turkey-us relations. *Political Psychology*, *31*(2), 275–303.

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., & Starz, J. (2018). Icews coded event data. *Havard Dataverse*, *V25*. Retrieved from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QI2T9A

Botto, K., & Jo, E. A. (2018). *The aftermath of the third inter-korean summit of 2018: Scenarios.* Retrieved from https://carnegieendowment.org/posts/2018/10/the-aftermath-of-the-third-inter-korean-summit-of-2018-scenarios?lang=en (Carnegie Endowment for International Peace, Accessed: 2025-04-15)

Boulding, K. E. (1956). *The image: Knowledge in life and society* (Vol. 47). University of Michigan press.

Brandt, P. T., Colaresi, M., & Freeman, J. R. (2008). The dynamics of reciprocity, accountability, and credibility. *Journal of Conflict Resolution*, *52*(3), 343–374.

Brookings Institution. (2018). *Around the halls: Brookings experts react to the trump–kim jong-un summit in singapore.* Retrieved from https://www.brookings.edu/articles/around-the-halls-brookings-experts-react-to-the-trump-kim-jong-un-summit-in-singapore/ (Accessed: April 17, 2025)

Cha, V. D. (2002). Korea's place in the axis. *Foreign Affairs*, 79–92.

Colaresi, M. P. (2004). When doves cry: International rivalry, unreciprocated cooperation, and leadership turnover. *American Journal of Political Science*, *48*(3), 555–570.

Converse, P. E. (1964). The nature of belief systems in mass publics (1964). *Critical review*, *18*(1-3), 1–74.

Cottam, R. W. (1977). *Foreign policy motivation: A general theory and a case study*. University of Pittsburgh Pre.

Davies, G. A., & Johns, R. (2016). The domestic consequences of international over-cooperation: An experimental study of microfoundations. *Conflict Management and Peace Science*, *33*(4), 343–360.

East Asia Institute. (2004). *Public opinion poll on the roh administration's first year (in korean).* Retrieved 2021-09-12, from http://www.eai.or.kr/data/bbs/kor_report/2009071414552332.pdf

Edgerton, J. F. (2024). Cooperative communities in the international system: Networked cooperation of similar states. *The Journal of Politics*, *86*(4), 1509–1523.

Eicher, V., Pratto, F., & Wilhelm, P. (2013). Value differentiation between enemies and allies: Value projection in national images. *Political Psychology*, *34*(1), 127–144.

Friedhoff, K. (2020). Americans positive on south korea despite trump's views on alliance. *The Chicago Council on Global Affairs*. Retrieved from https://globalaffairs.org/research/public-opinion-survey/americans-positive-south-korea-despite-trumps-views-alliance

Fu, R. T.-m., & Lee, E. I.-l. (2024). The impact of situational context and dispositions on costly signals of reassurance: Experimental evidence from taiwan. *Political Psychology*.

Gallup. (2018). *Daily opinion: Public opinion poll on the moon administration's first year (in korean).* Retrieved 2021-09-12, from https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=925

Gibler, D. M., Hutchison, M. L., & Miller, S. V. (2012). Individual identity attachments and international conflict: The importance of territorial threat. *Comparative Political Studies*, *45*(12), 1655–1683.

Goldstein, J. S., & Freeman, J. R. (1990). *Three-way street: Strategic reciprocity in world politics*. University of Chicago Press.

Goldstein, J. S., & Pevehouse, J. C. (1997). Reciprocity, bullying, and international cooperation: Time-series analysis of the bosnia conflict. *American Political Science Review*, 515–529.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, *21*(1), 107–112.

Heider, F. (1982). *The psychology of interpersonal relations*. Psychology Press.

Herrmann, R. K. (1985). *Perceptions and behavior in soviet foreign policy*. University of Pittsburgh Pre.

Herrmann, R. K. (2017). How attachments to the nation shape beliefs about the world: A theory of motivated reasoning. *International Organization*, S61–S84.

Herrmann, R. K., & Fischerkeller, M. P. (1995). Beyond the enemy image and spiral model: cognitive–strategic research after the cold war. *International organization*, *49*(3), 415–450.

Herrmann, R. K., Tetlock, P. E., & Visser, P. S. (1999). Mass public decisions to go to war: A cognitive-interactionist framework. *American Political Science Review*, 553–573.

Herrmann, R. K., Voss, J. F., Schooler, T. Y., & Ciarrochi, J. (1997). Images in international relations: An experimental test of cognitive schemata. *International Studies Quarterly*, *41*(3), 403–433.

Hogg, M. A. (1992). The social psychology of group cohesiveness: From attraction to social identity.

Hurwitz, J., & Peffley, M. (1990). Public images of the soviet union: The impact on foreign policy attitudes. *The Journal of Politics*, *52*(1), 3–28.

Jervis, R. (1976). *Perception and misperception in international politics: New edition*. Princeton University Press.

Jung, D.-j., Kim, S., Kim, H.-j., Na, Y., Moon, I.-c., Song, Y., … Lee, J. (2018). *2018 aunnual survey on south koreans' perception on north korea (in korean)*. Institute for Peace and Unification Studies, Seoul National University.

Kafura, C. (2022). American views of china remain at record lows. *The Chicago Council on Global Affairs*. Retrieved from https://globalaffairs.org/research/public-opinion-survey/american-views-china-remain-record-lows

Kaplowitz, N. (1990). National self-images, perception of enemies, and conflict strategies: Psychopolitical dimensions of international relations. *Political Psychology*, 39–82.

Katzenstein, P. J. (1996). *The culture of national security: Norms and identity in world politics*. Columbia University Press.

Kertzer, J. D., Rathbun, B. C., Rathbun, N. S., et al. (2020). The price of peace: Motivated reasoning and costly signaling in international relations. *International Organization*, *74*(1), 95–118.

Kertzer, J. D., & Zeitzoff, T. (2017). A bottom-up theory of public opinion about foreign policy. *American Journal of Political Science*, *61*(3), 543–558.

Kim, S. J. (2021). Quick on the draw: American negativity bias and costly signals in international relations. *Journal of Conflict Resolution*, 00220027211040104.

Kreps, S. E., Saunders, E. N., & Schultz, K. A. (2018). The ratification premium: Hawks, doves, and arms control. *World Politics*, *70*(4), 479–514.

Kriesberg, L. (2007). *Constructive conflicts: From escalation to resolution*. Rowman & Littlefield.

Kupchan, C. A. (2012). *How enemies become friends: The sources of stable peace*. Princeton University Press.

Lee, B. K. (2022). Triangles, major powers, and rivalry duration. *Journal of Conflict Resolution*, 00220027221134781.

Matsuzaki, T. (2020). *Nagisa: A japanese tokenizer for python.* https://pypi.org/project/nagisa/. (Accessed: 2025-04-13)

Mattes, M., & Weeks, J. L. (2019). Hawks, doves, and peace: an experimental approach. *American Journal of Political Science*, *63*(1), 53–66.

Mattes, M., & Weeks, J. L. (2022). Reacting to the olive branch: Hawks, doves, and public support for cooperation. *International Organization*, *76*(4), 957–976.

Menegazzi, S. (2018). *The european union can play a role on the korean peninsula.* Retrieved from https://eastasiaforum.org/2018/06/02/the-european-union-can-play-a-role-on-the-korean-peninsula/ (Accessed: April 17, 2025)

Migdal, J. S., et al. (2001). *State in society: Studying how states and societies transform and constitute one another*. Cambridge University Press.

Minhas, S., Hoff, P. D., & Ward, M. D. (2016). A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research*, *53*(3), 491–505.

Myrick, R., & Wang, C. (2023). Domestic polarization and international rivalry: How adversaries respond to america's partisan politics. *Journal of Politics*.

Onishi, N. (2007). *Korean summit results exceed low expectations.* Retrieved from https://www.nytimes.com/2007/10/05/world/asia/05korea.html (The New York Times, Accessed: 2025-04-15)

Park, E. L., & Cho, S. (2014). Konlpy: Korean natural language processing in python. In *Proceedings of the 26th annual conference on human & cognitive language technology* (Vol. 6, pp. 133–136).

Peffley, M., & Hurwitz, J. (1992). International events and foreign policy beliefs: Public response to changing soviet-us relations. *American Journal of Political Science*, 431–461.

Rasler, K., Thompson, W. R., & Ganguly, S. (2013). *How rivalries end*. University of Pennsylvania Press.

Rider, T. J., & Owsiak, A. P. (2021). *On dangerous ground: A theory of bargaining, border settlement, and rivalry*. Cambridge University Press.

Risse-Kappen, T. (2016). Collective identity in a democratic community. *Içinde Domestic Politics and Norm Diffusion in International Relations: Ideas do not float freely, Thomas Risse*, 78–113.

Schafer, M. (1997). Images and policy preferences. *Political Psychology*, *18*(4), 813–829.

Schultz, K. A. (2005). The politics of risking peace: Do hawks or doves deliver the olive branch? *International Organization*, 1–38.

Tajfel, H. (1981). *Human groups and social categories*. Cambridge university press Cambridge.

The Hankyoreh. (2007). *More than half of public disapproves of roh's diplomacy and security policies (in korean).* Retrieved from https://www.hani.co.kr/arti/politics/politics_general/192724.html (Accessed: 2025-04-15)

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237–252.

Vasquez, J. A. (1993). *The war puzzle* (Vol. 27). Cambridge University Press.

Ward, M. D., Beger, A., Cutler, J., Dickenson, M., Dorff, C., & Radford, B. (2013). Comparing gdelt and icews event data. *Analysis*, *21*(1), 267–297.

Wendt, A. (1994). Collective identity formation and the international state. *American political science review*, *88*(2), 384–396.

Wendt, A. (1999). *Social theory of international politics* (Vol. 67). Cambridge University Press.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge university press.