

# Does Neutrality Have a Price?

Experimental Evidence from Japan and the Philippines

Viet Hung Nguyen Cao 

Graduate School of Political Science, Waseda University

Strategic Studies Program, Center for Integrative and Development Studies, University of the Philippines, Diliman

[ncviethung@fuji.waseda.jp](mailto:ncviethung@fuji.waseda.jp)

Marvin Harmor Bernardo 

International Doctoral Program in Asia-Pacific Studies, National Chengchi University  
Strategic Studies Program, Center for Integrative and Development Studies, University of the Philippines, Diliman

[mhbernardo@up.edu.ph](mailto:mhbernardo@up.edu.ph)

2025-11-26

## Abstract

This study investigates democratic conflict behavior, through a survey experiment of type II audience cost in the Philippines and Japan in the event of Taiwan-Straight crisis. Looking at Filipino and Japanese public's reactions to hypothetical scenarios in which government leaders in each nation, respectively, initially declare they would not participate in a Taiwan crisis but would later change their position. The investigation aims to identify the political cost of justifying involvement in third-party conflicts in democratic states in East Asia. The project will contribute to the growing literature on audience cost, allies maintenance, and the relationship between democratic values and international crisis. It will also have implications for policy-making and crisis management in the event of a Taiwan Strait crisis.

**Keywords:** audience costs, Japan, the Philippines, Taiwan conflicts, survey experiment

**Words count:** 4494 words

# 1 Introduction

Audience costs theory has emerged as one of the most influential yet contentious frameworks in the study of international conflict. First systematically theorized by [Fearon \(1994\)](#), the concept posits that leaders who make public threats and subsequently back down face domestic political punishment from their constituencies. This anticipated punishment, the theory suggests, makes threats more credible and helps explain why democratic states may be more effective at deterrence and crisis bargaining. The appeal of audience costs lies in its elegant solution to the commitment problem in international relations: by tying their hands domestically, leaders can credibly signal resolve to adversaries, thereby reducing the likelihood of war through miscalculation ([Fearon, 1997](#); [Schultz, 1998](#)).

However, despite its theoretical prominence, audience costs theory has become increasingly controversial. A growing body of scholarship questions both the existence and magnitude of audience costs, with empirical studies producing mixed and often contradictory findings ([Snyder and Borghard, 2011](#); [Trachtenberg, 2012](#); [Gartzke and Lupu, 2012](#)). Scholars have challenged the original formulation on multiple fronts, including the conditions under which audience costs emerge, their relative size compared to other political considerations, and whether they operate symmetrically across regime types ([Weeks, 2008](#); [Weiss, 2013](#)). The debate has intensified as researchers have sought to test the theory's microfoundations through experimental and observational methods, revealing a more complex empirical landscape than early theoretical models suggested ([Tomz, 2007](#); [Kertzer and Brutger, 2016](#)).

One particularly important dimension of this debate concerns what has been termed the “inconsistency mechanism” underlying audience costs. As [Levy et al. \(2015\)](#) have argued, if domestic audiences punish leaders for inconsistency between words and deeds, then the leaders should face costs not only for *backing down* from conflicts they threatened to engage in, but also for *backing into* conflicts they initially pledged to avoid. This latter phenomenon, called Type II audience costs ([Quek, 2017](#)), represents a theoretically significant extension of the original framework. Type II audience costs would suggest that public commitments to peace or non-intervention also create domestic constraints, potentially limiting leaders' flexibility in responding to international crises ([Kertzer and Brutger, 2016](#); [Quek, 2017](#)). The empirical evidence for Type II audience costs, however, remains decidedly mixed. While some experimental studies have found evidence that leaders face punishment for backing into conflicts ([Levy et al., 2015](#)), others have found weaker or null effects ([Takei and Paolino, 2023](#)). Critically, the vast majority of this research has been conducted in the United States, examining American public opinion toward hypothetical foreign policy scenarios. Whether the findings from American samples generalize to other democracies remains an open question. The external validity of Type II audience costs theory thus depends on expanding empirical investigation beyond the American context to include other countries.

This research note addresses this gap by examining Type II audience costs in two key US allies in the Asia-Pacific region: Japan and the Philippines. We conducted parallel survey experiments in both countries in September 2025, employing scenarios designed to test whether publics punish leaders for abandoning neutrality and entering conflicts at the behest of the United States. Our findings reveal a nuanced pattern that challenges straightforward applications of Type II audience costs theory to allied contexts. Contrary

to the prediction that leaders face punishment for backing into conflicts, we find that Type II audience costs effectively disappear in Japan and the Philippines when backing into conflicts occurs within the context of assisting the US. However, this finding emerges not because respondents do not value consistency in leaders' words and deeds, but rather because they assign greater importance to maintaining the alliance relationship with the United States than to preserving their country's initial position of neutrality.

This finding carries important implications for both theory and policy. Theoretically, it suggests that the inconsistency mechanism underlying audience costs may be more context-dependent than previously recognized, particularly in alliance relationships characterized by asymmetric power. For junior allies dependent on a superpower patron for security, alliance credibility may override concerns about domestic political consistency regarding specific conflicts. Empirically, our study makes two key contributions. First, it provides much-needed evidence on audience costs dynamics outside the United States, examining two democracies with distinct strategic situations. Second, it offers new insight into the specific boundary conditions under which Type II audience costs emerge or fail to emerge, highlighting the role of alliance relationships in shaping public attitudes toward conflict escalation. Practically, these findings hold significant relevance for policymakers in Washington, Tokyo, and Manila as they prepare for potential contingencies in the Taiwan Strait and broader Indo-Pacific region. Understanding how publics in allied countries respond to scenarios involving entrapment is crucial for assessing the political sustainability of coordinated military action and for managing alliance relationships during crises.

## 2 Motivation and Context

There are several reasons why we chose Japan and the Philippines to conduct our survey experiments. First, we are interested in understanding if being entrapped in an unwanted conflict would influence how the public judge the inconsistency of their leaders. Entrapment, the fear that security guarantee of the alliance will embolden a party to take aggressive action, dragging the other party into an unwanted conflict, is one of the two main dilemmas of alliance politics ([Snyder, 1984](#)). Most studies have focused on how the junior partners of the US might be emboldened by the alliance, and would drag the US into war against their interests ([Benson, 2012](#); [Beckley, 2015](#)). An emerging body of literature examined if the public in the junior partners will indeed be emboldened by US signals, with mostly mixed or even negative results ([Sukin, 2020](#); [Kudo and Nguyen Cao, 2025](#)). However, to our knowledge, our study will be the first one to examine public opinions of US junior allies in the scenario where they are the parties being entrapped in unwanted conflicts. By examining Type II audience costs in the context of alliance relationships, we can assess whether the inconsistency of abandoning neutrality is evaluated differently when it occurs in service of alliance obligations rather than through independent decision-making. This focus on entrapment dynamics allows us to test whether the structural asymmetry inherent in relationships between a superpower and its junior allies conditions the operation of audience costs mechanisms.

Second, both Japan and the Philippines are established democracies. As argued by [Fearon \(1994\)](#), democratic leaders have an advantage in signaling audience cost. Therefore, if we want to verify the external validity of Type II audience cost beyond the US, both of

these countries are ideal cases. If leaders in Tokyo and Manila face the same domestic backlash for backing into a conflict like those in Washington, we can be more confident in the existence and mechanism of this type of audience cost. However, even if we fail to replicate the findings of previous works, we learn more about the boundaries conditions of this theory, as well as how other factors, such as perceived alliance commitments, might influence citizens' support for intervention in foreign conflicts.

Another theoretical motivation for us is to verify the mechanism of audience cost mitigation. If Type II audience cost exists and shares the same mechanism with Type I audience cost, as argued in previous works, then this domestic backlash can also be mitigated the same way. Following [Levendusky and Horowitz \(2012\)](#), we examine if providing justifications for the inconsistency between words and deeds can help leaders reduce the domestic hurdles for backing into conflicts. Understanding mitigation strategies is crucial both theoretically, as it illuminates the cognitive processes underlying audience judgments of consistency, and practically, as it informs how leaders can communicate foreign policy reversals to domestic audiences.

We also selected Japan and the Philippines because they are the ideal cases to deploy a realistic, highly salient scenario. As mentioned before, the previous works on Type II audience costs were done with samples of US citizens. The scenarios were turned into ones about a third party dispute, in which the US audience has no salient connection to. The use of anonymous conflict parties in previous studies also voided the respondents of the consideration over the important characteristics of the parties, such as regime type, and the stakes of the conflict. These considerations might shape the foreign policy preferences of the public in the real world. Our experiments use a realistic Taiwan contingency scenario, which is highly salient to both Japanese and Philippine publics given their geographic proximity, economic ties, and security interests in cross-Strait stability. The use of a real-world case offers several advantages. It increases external validity by embedding the inconsistency dilemma within an actual geopolitical context about which respondents hold prior beliefs and preferences. It allows us to examine how substantive features of the conflict, including the identities of the parties involved and the stakes of the conflict, influence the magnitude of audience costs.

### 3 Hypotheses

Based on these motivations, we propose three hypotheses with this research note. The first hypothesis is consistent with the logic of Type II audience cost. The leaders who break the promise of neutrality would have lower support than those who keep it.

#### Hypothesis 1:

The public's support for leaders would be lower when leaders decide to participate in conflict despite previously promised not to, compared to when they keep their promise not to participate.

Regarding mitigation strategies, this study follows previous works, such as [Levendusky and Horowitz \(2012\)](#) or [Quek and Johnston \(2018\)](#), in giving the respondents new information. We are interested primarily in the effects of two factors: shared democratic identity and national security. We deliberately chose these two treatments for several reasons. First, the works on democratic peace has established that democracies not only

do not fight wars with each other (Doyle, 1986; Maoz and Russett, 1993), they might even perceive other democracies as less threatening (Tomz and Weeks, 2013). Chu et al. (2021) showed that perceived shared democratic values increase the public's perceived in-group cohesion, boosting their approval to protect another democracy. For these reasons, we propose that if the leaders of Japan or the Philippines wish to mitigate Type II audience costs in a Taiwan contingency, reminding the public that Taiwan is also a democracy might be a good rhetorical strategy.

### Hypothesis 2:

The public's support for leaders would be higher when leaders decide to justify their decision based on the need to help another democracy, compared to when they do not give any justification.

Similarly, we predict that the leaders can also justify their decision to back into a conflict from the perspective of national security. In the specific context of a Taiwan contingency, national security arguments carry particular weight for both Japan and the Philippines given their geographic proximity to potential conflict, their economic dependence on stable maritime routes through the Taiwan Strait, and their own territorial disputes with China that make them stakeholders in the broader regional balance of power.

### Hypothesis 3

The public's support for leaders would be higher when leaders decide to justify their decision based on national security, compared to when they do not give any justification.

## 4 Experimental design

We conducted two parallel survey experiments in Japan and the Philippines in September 2025.<sup>1</sup> In both experiments, we first presented all respondents with a vignette regarding a hypothetical scenario in which China is ramping up their military activities around the Taiwan Strait. Respondents were informed that a crisis might be imminent, and that the leader of their respective country announced that they are calling on both sides to resolve the conflict peacefully, and that their country is not taking any side and would try to stay out of the conflict. After reading this common vignette, respondents were randomized into 4 experimental groups, 1 control group and 3 treatment groups. All groups were informed that China decided to launch an invasion of Taiwan, and that the US mobilized their forces in the region, as well as calling on their countries to join forces. In the control group, respondents were informed that the leader decided to keep their previous promise and refused to help the US. In the first treatment group, respondents were presented with a statement from their leader saying that they agreed to help the US. In the second and third treatment groups, in addition to the statement similar to the first treatment group, respondents were given additional justifications of helping Taiwan, a fellow democracy, and helping Taiwan based on national security.<sup>2</sup>

All respondents were then asked to indicate their approval or disapproval for the action

---

<sup>1</sup>The experiments were pre-registered before the collection of data. The pre-registration can be found here ([https://osf.io/2xne8/?view\\_only=a0e509e09e044b078f922d46316008cf](https://osf.io/2xne8/?view_only=a0e509e09e044b078f922d46316008cf)).

<sup>2</sup>The exact wordings of the vignette can be found in Appendix A.1.

of the leader. The answers were recorded on a 5-point Likert-scale, from “Strongly disapprove” (numerical value of -2) to “Strongly approve (numerical value of 2). They were also asked to indicate the reason for approval/disapproval in a short sentence, from 10-20 words. We also sought to understand the extent to which the respondents think the action weaken or strengthen the alliance between their countries with the US. Respondents were also asked to point out the responses to the conflict that they want their government to take.

We fielded both experiments from September 24 to 26, 2025, through PureSpectrum. We collected a Japanese sample of 1,198 respondents and a Filipino sample of 1,200 respondents. While we applied quota sampling to ensure the collected samples were as similar to the respective censuses as possible, the collected samples still differed slightly from the censuses. However, the differences were negligible and did not impact the analyses.<sup>3</sup>

## 5 Results

### 5.1 Overall results

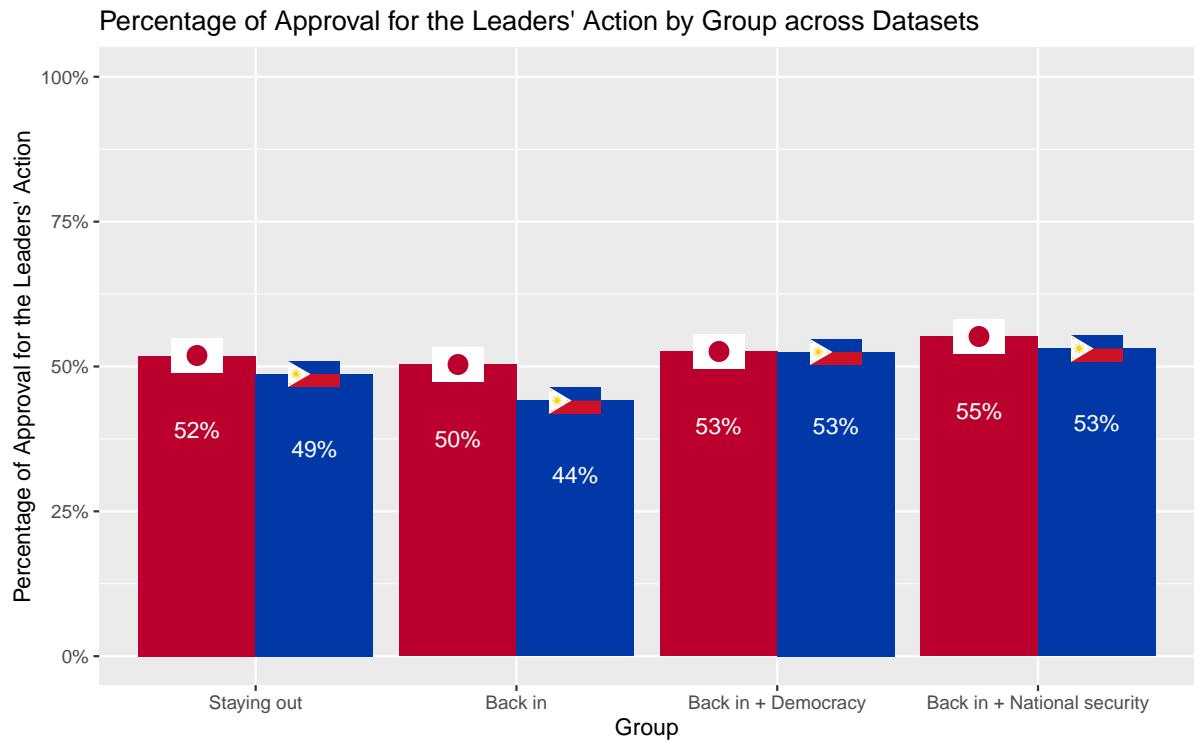


Figure 1: Approval of the leaders’ actions

The collected data were analyzed with ANOVA and adjust p-values with TukeyHSD method for pairwise multiple means comparisons. The results were also confirmed with

---

<sup>3</sup>Detailed on the samples, sampling methods can be found in Appendix B. We also conducted balance checks to determine the distribution of treatments among the groups and found no imbalance. The full balance checks can be found in Appendix C.

both regression analyses and equivalence tests.<sup>4</sup> Figure 1 summarizes the approval rates for the leaders' action in both countries. Overall, we do not find support for any of the proposed hypotheses. The differences between the approval rate of the control group and all treatment groups are negligible.

Overall, the results of the experiment in Japan showed that the Japanese leaders would not face any domestic challenges for abandoning a promise of neutrality if Taiwan is invaded by China. The difference between keeping the promise (control group) and joining the US in defending Taiwan (treatment group 1) is largely statistically insignificant (diff-in-means = -0.046, 95% CI [-0.249, 0.157],  $p = 0.938$ ). We also do not observe any significant influence of the justifications, democracy (diff-in-means = 0.085, 95% CI [-0.117, 0.288],  $p = 0.697$ ) or national security (diff-in-means = 0.119, 95% CI [-0.084, 0.322],  $p = 0.435$ ). These null results provide no empirical evidence to support any of the proposed hypotheses.

Similarly, we also do not find support for any of the proposed hypotheses in the Philippines. The difference between keeping the promise of staying out (control group) and backing in (treatment 1) is also statistically insignificant (diff-in-means = -0.117, 95% CI [-0.356, 0.121],  $p = 0.585$ ). If we are to interpret this null result as the lack of domestic audience cost for backing in, it is by no surprise that none of the justifications for backing in influence the approval of the public. Both emphasizing shared democratic regimes (diff-in-means = 0.157, 95% CI [-0.084, 0.398],  $p = 0.335$ ) and national security (diff-in-means = 0.226, 95% CI [-0.013, 0.466],  $p = 0.072$ ) yielded null results.

## 5.2 The effects of alliance with the US

In both experiments, we also measured how the respondents think the action of their leaders would impact their country's alliance with the US. We believe that this is an important moderator influencing the approval for staying out or backing in, for two main reasons. First of all, the public in these two countries might view a refusal to heed the call of the US negatively, because they might be concerned that given the refusal, the alliance between their country and the US would be weakened. Vice versa, if the respondents perceive that helping the US with Taiwan might boost the alliance between their own country and the US, they might not judge the leader's decision to abandon the promise of neutrality harshly. The other reason why this variable can moderate the approval for backing in or staying out is because the respondents might perceive backing in as unavoidable when their country is called upon by the US. This perception might provide an explanation to the difference in type II audience costs between the public in these junior partners of the US, and the public in the US itself.

---

<sup>4</sup>The results of the regressions can be found in Appendix A.4, the results of the equivalence tests can be found in Appendix E.

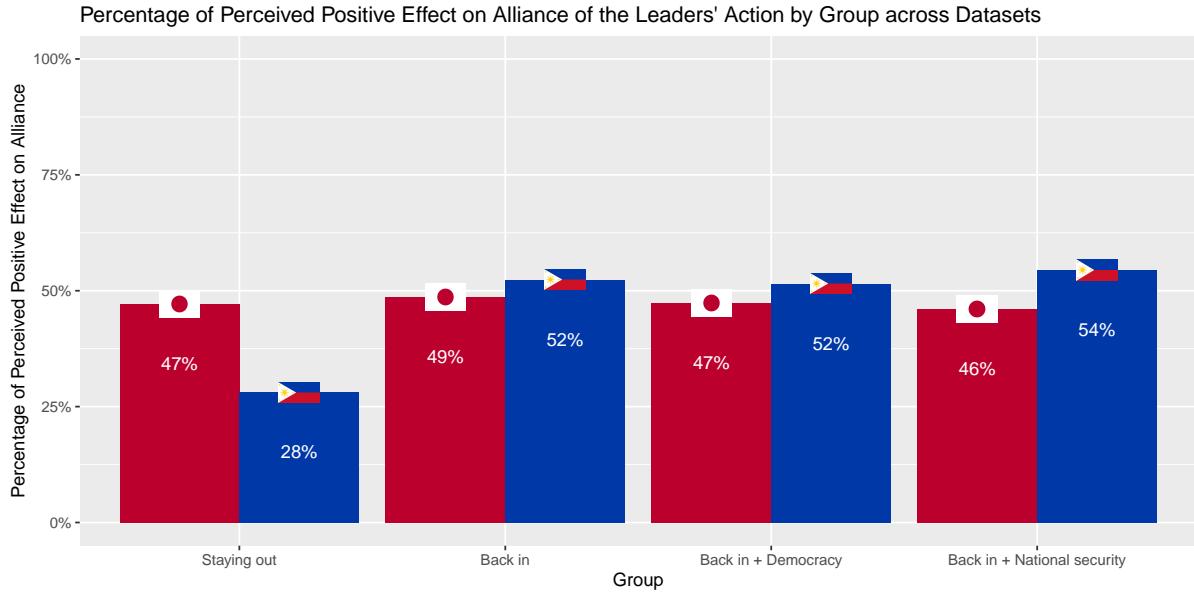


Figure 2: Perceived impact of leaders' action on alliance

Figure 2 summarized the perceived impact of the leaders' action on the alliance with the US. We observed somewhat diverged perceptions between Japanese and Filipino public. While the Japanese public seems confident in their alliance with the US, and regardless of whether Japan helps the US in Taiwan or not, they think the alliance would still be intact. On the other hand, the Filipino public showed more insecurities. If the Filipino leaders do not heed the call of the US to help in Taiwan, the public believes that this action would weaken the alliance, and vice versa.

This divergence among the public about the impacts on their respective alliances is also reflected in the moderating effect of the alliance factor on the public's approval/disapproval of their leaders' action. For the Japanese public, our results show that there is a weak negative backlash (significant at the 90% level) for the leader, if the public perceives that reneging on the promise of neutrality poses no consequence for the alliance with the US. However, if the respondents believe that helping the US in Taiwan will strengthen the alliance, this negative effect disappeared. In fact, even when the leader does not keep the promise of neutrality, if the public believes that joining in the conflict will help the alliance, they will support that decision.

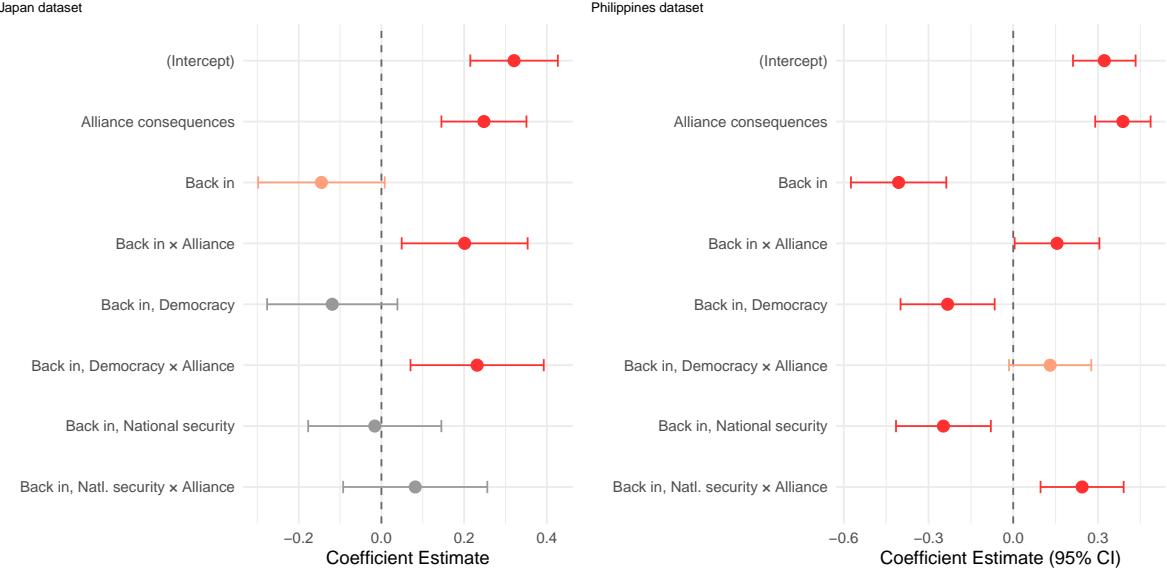


Figure 3: Perceived impact of leaders' action on alliance

This moderating effect is even stronger in the Philippines. As can be seen from Figure 3, when the action of the leaders is perceived as not having consequence for the alliance, the leader would face significant backlash for breaking a promise of neutrality. The negative effects are statistically significant across all treatment groups. However, if going back on the promise means a stronger alliance with the US, the leader would actually receive public support. This positive effect is also statistically significant across the treatment groups.

## 6 Discussion

Our results provide more nuanced insight into Type II audience costs, especially in the junior partners of the US in the Asia-Pacific. On the one hand, our diff-in-means results of the experiments showed that the leaders in these countries do not face audience costs for breaking a promise of staying out of the conflict in Taiwan. This is contrary to the predictions based on previous results of Type II audience costs by Levy et al. (2015) and Quek (2017), and support the results by Takei and Paolino (2023) which questioned both the existence and magnitude of the audience costs for backing in.

On the other hand, our results showed that the lack of direct audience costs for breaking the promise might be due to the strong effect of the alliance with the US in these countries. The analysis of our consequences for alliance moderator showed that if the respondents think that breaking the promise would entail no impact for the alliance, the leaders would at least face some negative consequences, which is consistent with the Type II audience costs predictions. However, because of the anticipated consequences for the alliance, this negative effect is reversed, because the respondents believe backing in might help strengthening the alliance, and hence they would support the leader even when it means breaking the promise. The rather surprising results here might indicate that respondents in these countries care less about the consistency in the words and actions of the leaders, and judge them more on the nature of the actions that they take in crises.

Taken together, the moderating effects of the alliance factor in both countries showed that for these junior partners of the US, the domestic audience cares less about the consistency in the words and actions of the leaders. Even when dragged into an unwanted conflict by their senior partner, they would support the leader breaking a promise if doing so would strengthen the alliance with the US. In other words, the leaders in the junior partners do not need to worry about any domestic audience costs for being involved in a conflict which they promised to stay out of.

## 7 Conclusion

This research note set out to examine whether Type II audience costs, the domestic political punishment leaders face for backing into conflicts they initially pledged to avoid, operate in two key US allies in the Asia-Pacific: Japan and the Philippines. Our parallel survey experiments, conducted in September 2025 using realistic Taiwan contingency scenarios, produced findings that simultaneously challenge and refine existing theories of audience costs in alliance contexts.

Our results reveal that leaders in Japan and the Philippines do not face significant Type II audience costs for abandoning neutrality and supporting US military intervention in a Taiwan crisis. The direct comparison between maintaining neutrality and backing into the conflict yielded no statistically significant differences in public approval across either country. At first glance, these null findings might suggest that Type II audience costs simply do not exist outside the United States, or that the inconsistency mechanism underlying audience costs theory does not operate uniformly across democratic contexts. However, our investigation of moderating factors reveals a more nuanced and theoretically important story.

The critical insight from our study emerges from analyzing how perceptions of alliance consequences shape public attitudes toward leader inconsistency. When respondents believed that abandoning neutrality would strengthen the alliance with the United States, or conversely, that maintaining neutrality would weaken it, the expected Type II audience costs disappeared entirely and even reversed direction. In the Philippines, this pattern was particularly pronounced: leaders faced significant domestic backlash for breaking neutrality promises when respondents perceived no alliance benefits, but received substantial public support when backing into the conflict was viewed as strengthening US-Philippine ties. In Japan, the pattern was somewhat weaker but still evident, with alliance considerations overriding concerns about consistency.

Methodologically, our study addresses several limitations of previous research. By examining two democracies outside the United States with distinct strategic situations, we enhance the external validity of audience costs theory and demonstrate that findings from American samples do not necessarily generalize to other democratic contexts. By employing realistic scenarios involving Taiwan rather than anonymous conflicts, we capture the complex attitudes and strategic considerations that shape actual foreign policy preferences. By directly measuring perceptions of alliance consequences, we identify a crucial moderating variable that may account for inconsistent findings in earlier experimental work.

The practical implications of our findings merit serious attention from policymakers in

Washington, Tokyo, and Manila. As regional tensions over Taiwan continue to escalate and allied nations engage in contingency planning, understanding public attitudes toward intervention becomes crucial for assessing the political sustainability of coordinated military action. Our results suggest that leaders in Japan and the Philippines possess greater flexibility than Type II audience costs theory would predict when it comes to reversing initial positions of neutrality—but only insofar as publics view such reversals as serving alliance interests. This finding has two important implications for crisis management. On one hand, it suggests that fears of domestic opposition to supporting the United States in a Taiwan contingency may be overstated, at least in these two countries. On the other hand, it underscores the importance of how leaders frame involvement: emphasizing alliance benefits appears more politically effective than appealing to shared democratic values or even national security considerations in isolation.

## References

- Beckley, Michael (2015), “The myth of entangling alliances.” *International Security*, 39, 7–48.
- Benson, Brett V. (2012), *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. Cambridge University Press.
- Chu, Jonathan A., Jiyoung Ko, and Adam Liu (2021), “Commanding support: Values and interests in the rhetoric of alliance politics.” *International Interactions*, 47, 477–503.
- Doyle, Michael W (1986), “Liberalism and world politics.” *The American Political Science Review*, 80, 1151–1169.
- Fearon, James D. (1994), “Domestic Political Audiences and the Escalation of International Disputes.” *American Political Science Review*, 88, 577–592.
- Fearon, James D. (1997), “Signaling foreign policy interests: Tying hands versus sinking costs.” *Journal of Conflict Resolution*, 41, 68–90.
- Gartzke, Erik and Yonatan Lupu (2012), “Still looking for audience costs.” *Security Studies*, 21, 391–397.
- Kertzer, Joshua D. and Ryan Brutger (2016), “Decomposing audience costs: Bringing the audience back into audience cost theory.” *American Journal of Political Science*, 60, 234–249.
- Kudo, Yasuki and Viet Hung Nguyen Cao (2025), “Does the us really embolden its allies? evidence from a survey experiment in japan.” *Journal of Conflict Resolution*, 00220027251360557.
- Levendusky, Matthew S. and Michael C. Horowitz (2012), “When backing down is the right decision: Partisanship, new information, and audience costs.” *The Journal of Politics*, 74, 323–338.
- Levy, Jack S., Michael K. McKoy, Paul Poast, and Geoffrey P.R. Wallace (2015), “Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory.” *American Journal of Political Science*, 59, 988–1001.
- Maoz, Zeev and Bruce Russett (1993), “Normative and structural causes of democratic peace, 1946–1986.” *The American Political Science Review*, 87, 624–638.
- Quek, Kai (2017), “Type II Audience Costs.” *The Journal of Politics*, 79, 1438–1443.
- Quek, Kai and Alastair Iain Johnston (2018), “Can China Back Down? Crisis De-escalation in the Shadow of Popular Opposition.” *International Security*, 42, 7–36.
- Schultz, Kenneth A. (1998), “Domestic opposition and signaling in international crises.” *American Political Science Review*, 92, 829–844.
- Snyder, Glenn H. (1984), “The security dilemma in alliance politics.” *World Politics*, 36, 461–495.
- Snyder, Jack and Erica D. Borghard (2011), “The cost of empty threats: A penny, not a pound.” *American Political Science Review*, 105, 437–456.

Sukin, Lauren (2020), “Credible nuclear security commitments can backfire: Explaining domestic support for nuclear weapons acquisition in south korea.” *Journal of Conflict Resolution*, 64, 1011–1042.

Takei, Makito and Philip Paolino (2023), “Backing Out but Backing In Audience Costs? A Replication of Levy et al. (2015).” *Foreign Policy Analysis*, 19, orad008.

Tomz, Michael (2007), “Domestic Audience Costs in International Relations: An Experimental Approach.” *International Organization*, 61.

Tomz, Michael R. and Jessica L. P. Weeks (2013), “Public opinion and the democratic peace.” *American Political Science Review*, 107, 849–865.

Trachtenberg, Marc (2012), “Audience costs: An historical analysis.” *Security Studies*, 21, 3–42.

Weeks, Jessica L. (2008), “Autocratic audience costs: Regime type and signaling resolve.” *International Organization*, 62.

Weiss, Jessica Chen (2013), “Authoritarian signaling, mass audiences, and nationalist protest in china.” *International Organization*, 67, 1–35.

# A Appendix

## A.1 Experimental vignettes

### A.1.1 Japan vignettes

It is the year 20xx. China has increased the frequency of military activities around the Taiwan Strait. Experts determined that a potential crisis would break out imminently. When asked by a reporter on which side of the conflict would Japan choose, the Prime Minister said: "Japan calls on both sides to resolve the conflict peacefully. We are not taking any side and would try to stay out of the conflict".

20XX年、中国は台湾海峡周辺での軍事活動を活発化させており、専門家は危機の勃発が差し迫っていると分析している。記者から本件に関して日本の立場を問われた際、内閣総理大臣は次のように述べた。

「日本は双方に対し、平和的な手段による紛争の解決を強く求める。我が国はいかなる陣営にも与せず、本件への関与を極力回避する方針である。」

#### *Control*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on Japan to join forces. The Prime Minister refused to support the US.

数か月が経過した。中国は台湾本島への侵攻を決断し、米国は自国の海軍部隊を当該地域に展開し、台湾海峡へと派遣した。米国は日本に対し協力を要請したが、内閣総理大臣はこれを支持しない意向を表明した。

#### *Treatment 1*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on Japan to join forces. The Prime Minister agreed to support the US.

数か月が経過した。中国は台湾本島への侵攻を決断し、米国は自国の海軍部隊を当該地域に展開し、台湾海峡へと派遣した。米国は日本に対し協力を要請し、内閣総理大臣はこれを支持する意向を表明した。

#### *Treatment 2*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on Japan to join forces. The Prime Minister agreed to support the US. In a statement, the Prime Minister explained that it is necessary to help Taiwan, a democracy like Japan.

数か月が経過した。中国は台湾本島への侵攻を決断し、米国は自国の海軍部隊を当該地域に展開し、台湾海峡へと派遣した。米国は日本に対し協力を要請し、内閣総理大臣はこれを支持する意向を表明した。声明の中で、総理大臣は、日本と同じ民主主義国家である台湾を支援することが必要であると強調した。

#### *Treatment 3*

A few months have passed. China decided to launch an invasion on the main island of

Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on Japan to join forces. The Prime Minister agreed to support the US. In a statement, the Prime Minister explained that it is necessary to help Taiwan from the perspective of national security.

数か月が経過した。中国は台湾本島への侵攻を決断し、米国は自国の海軍部隊を当該地域に展開し、台湾海峡へと派遣した。米国は日本に対し協力を要請し、内閣総理大臣はこれを支持する意向を表明した。声明の中で、総理大臣は、日本の国家安全保障の観点から台湾を支援することが必要であると強調した。

### A.1.2 Philippines vignettes

It is the year 20xx. China has increased the frequency of military activities around the Taiwan Strait. Experts determined that a potential crisis would break out imminently. When asked by a reporter on which side of the conflict would the Philippines choose, the President said: "The Philippines calls on both sides to resolve the conflict peacefully. We are not taking any side and would try to stay out of the conflict".

Ito ay ang taong 20xx. Pinapataas ng Tsina ang dalas ng mga gawaing militar sa paligid ng Taiwan Strait. Ayon sa mga ekspersto, isang potensyal na krisis ang maaaring biglang umusbong. Nang tanungin ng isang reporter kung aling panig ng hidwaan ang papanigan ng Pilipinas, sabi ng Pangulo: "Nanawagan ang Pilipinas sa magkabilang panig mapayapang lutasin ang kanilang alitan. Wala kaming kinakampihan at susubukan namin lumayo sa tunggalian".

#### *Control*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on the Philippines to join forces. The President refused to support the US.

Lumipas ang ilang buwan. Nagpasya ang China na maglunsad ng pagsalakay sa pangunahing isla ng Taiwan. Nagpasya ang US na pakilusin ang kanilang hukbong pandagat sa rehiyon patungong Taiwan Strait. Nanawagan sila sa Pilipinas para magsanib pwersa. Tumanggi ang Pangulo na suportahan ang US.

#### *Treatment 1*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on the Philippines to join forces. The President agreed to support the US.

Lumipas ang ilang buwan. Nagpasya ang China na maglunsad ng pagsalakay sa pangunahing isla ng Taiwan. Nagpasya ang US na pakilusin ang kanilang hukbong pandagat sa rehiyon patungong Taiwan Strait. Nanawagan sila sa Pilipinas para magsanib pwersa. Pumayag ang Pangulo na suportahan ang US.

#### *Treatment 2*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on the Philippines to join forces. The President agreed to support the US.

In a statement, the President explained that it is necessary to help Taiwan, a democracy like the Philippines.

Lumipas ang ilang buwan. Nagpasya ang China na maglunsad ng pagsalakay sa pangunahing isla ng Taiwan. Nagpasya ang US na pakilusin ang kanilang hukbong pandagat sa rehiyon patungong Taiwan Strait. Nanawagan sila sa Pilipinas para magsanib pwersa. Pumayag ang Pangulo na suportahan ang US. Sa isang pahayag, ipinaliwanag ng Pangulo na kailangang tulungan ang Taiwan, bilang isang demokrasya tulad ng Pilipinas.

### *Treatment 3*

A few months have passed. China decided to launch an invasion on the main island of Taiwan. The US decided to mobilize their naval forces in the region to the Taiwan Strait. They called on the Philippines to join forces. The President agreed to support the US. In a statement, the President explained that it is necessary to help Taiwan from the perspective of national security.

Lumipas ang ilang buwan. Nagpasya ang China na maglunsad ng pagsalakay sa pangunahing isla ng Taiwan. Nagpasya ang US na pakilusin ang kanilang hukbong pandagat sa rehiyon patungong Taiwan Strait. Nanawagan sila sa Pilipinas para magsanib pwersa. Pumayag ang Pangulo na suportahan ang US. Sa isang pahayag, ipinaliwanag ng Pangulo na kailangang tulungan ang Taiwan mula sa pananaw ng pambansang seguridad.

## **A.2 Sample information**

## **A.3 Balance checks**

## A.4 Regression analyses

### A.4.1 Japan

Table A1: OLS and Logit models - Japan

	OLS	OLS	Logit	Logit
Back In	-0.046 (0.079)	-0.090 (0.074)	-0.060 (0.164)	-0.156 (0.177)
Back In + Democracy	0.040 (0.079)	-0.046 (0.073)	0.032 (0.163)	-0.092 (0.176)
Back In + Nat Security	0.073 (0.079)	-0.021 (0.074)	0.136 (0.164)	0.015 (0.178)
Impact on Alliance		0.364*** (0.030)		0.608*** (0.077)
Nationalistic sentiment		0.124*** (0.034)		0.213* (0.083)
Perception of threat - China		-0.100 (0.111)		-0.354 (0.269)
Feeling toward China		0.004*** (0.001)		-0.002 (0.003)
Feeling toward Taiwan		0.001 (0.001)		0.012*** (0.003)
Internationalism		-0.006 (0.067)		-0.263 (0.162)
Num. Obs.	1198	1198	1198	1198

• p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

### A.4.2 Philippines

Table A2: OLS and Logit models - Philippines

	OLS	OLS	Logit	Logit
Back In	-0.117 (0.093)	-0.390*** (0.082)	-0.184 (0.164)	-0.771*** (0.192)
Back In + Democracy	0.040 (0.092)	-0.213** (0.081)	0.153 (0.163)	-0.332+ (0.189)
Back In + Nat Security	0.109 (0.092)	-0.201* (0.081)	0.177 (0.162)	-0.405* (0.189)
Impact on Alliance		0.460*** (0.028)		0.894*** (0.073)
Nationalistic sentiment		0.181*** (0.032)		0.258*** (0.076)
Perception of threat - China		0.044 (0.058)		0.076 (0.133)
Feeling toward China		-0.002 (0.001)		-0.004 (0.002)
Feeling toward Taiwan		0.003+ (0.002)		0.007+ (0.004)
Internationalism		-0.060 (0.066)		-0.242 (0.155)
Num.Obs.	1200	1200	1200	1200

• p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## A.5 Equivalence tests