**British Journal of
Political Science**

LETTER

Untying Hands: De-escalation, Reputation, and Dynamic Audience Costs

Kai Quek [iD]

Department of Politics and Public Administration, University of Hong Kong, Hong Kong
Corresponding author. Email: quek@hku.hk

Abstract
Two states in a dispute refuse to back down. One ties its own hands to strengthen its stand and gain advantage; the other tries to untie the tied hands to preempt disadvantage. Tying hands is a well-studied strategy, but it tells only part of the story, and the response strategy of untying hands remains unexplored. Can a state untie the tied hands of its opponent to give freedom back to its opponent—the freedom to concede? I identify three strategies of untying hands: counterthreat, reassurance, and normative framing. I show experimentally that these strategies can reduce the public costs of backing down and the perceived reputational damage from backing down. Tied hands and audience costs are not static and immutable, but dynamic and malleable by the other side.

How can a state communicate its resolve credibly and commit itself to a future action? A standard answer is tying hands,[1] a signaling mechanism that applies across many domains, from crisis management and terrorism to international cooperation and central bank independence.[2] Signalers tie their hands by issuing a signal that binds them to a future cost if they renege (Fearon 1997). An example of such costs is audience costs—the political losses a leader suffers for making a threat and then backing down (Fearon 1994).

A large literature in international relations (IR) has offered important insights into tying hands and audience costs. However, most studies assume an opponent that takes the tied hands and audience costs as given. Once the signaler ties its hands in a crisis, the opponent takes the tied hands as tied and is forced to face a strategic situation where the signaler now has less incentive to back down. In short, the opponent has become constrained by the signaler constraining itself.

But are tied hands truly tied? Or, can the opponent untie the tied hands to give freedom back to the signaler—the freedom to concede? These questions have important theoretical and policy implications, but they have not been directly addressed. Recent work has considered the agency of a signaler attempting to loosen its tied hands *ex post*, but the agency of the opponent remains unexplored. I focus on the opponent's agency and efficacy here.

---

[1]Recent work suggests another answer in the form of "reducible costs"—costs that were paid in the past but can be offset in the future depending on the signaler's action (Quek 2021a; see also Altman and Quek 2021). A tied-hands signal involves *ex post* costs; a reducible-cost signal involves *ex ante* costs.

[2]See, e.g., Conrad, Conrad, and Young (2014), Fischer (1995), Leeds (1999), Uzonyi, Souva, and Golder (2012), and Yoder (2019).

Contrary to the conventional depiction of a static international audience, the story here is a dynamic one. I propose that opponents do not necessarily take the signaler's tied hands as given; crisis diplomacy may involve an iterative process, with the signaler tying hands and the opponent untying hands; and the opponent can successfully untie the signaler's hands with some specific strategies. If these premises are correct, then tied hands and audience costs are not static—they are dynamic.

I proceed as follows. First, I develop the theoretical premises and propose three strategies of untying hands. Then, I describe the design and report the results of an experiment, showing how particular actions by the opponent can change the domestic costs and reputational impact of backing down. Finally, I discuss the implications of my findings, highlighting how they might enrich the strategic dynamics of signaling and credibility.

Untying Tied Hands

Leaders may tie their hands to commit themselves to a future action and communicate their resolve. In crisis diplomacy, tying hands is to take "an action that increases the costs of backing down if the would-be challenger actually challenges but otherwise entails no costs if no challenge materializes" (Fearon 1997, 70). The canonical example of tying hands is by generating audience costs (for a review, see Chan, Liu, and Quek 2021a).

To fulfill its strategic purpose, tied hands must remain tied. Tying hands works through an "irreversible sacrifice of freedom of choice" (Schelling 1960, 22). Most works on tying hands and audience costs have therefore taken tied hands and audience costs as given. There are some exceptions. Schultz (2001a) highlights how the credibility of audience-cost signals can depend on the reactions of opposition parties in a democratic state. Kurizaki (2007) suggests that state leaders may opt for secret diplomacy when their public threats would generate audience costs not only for themselves, but also for their opponents. Wu (2020) assumes leaders may manipulate their public threats to signal their domestic audience's preference for peace. Recent experimental work shows that leaders can use *ex post* actions to reduce their domestic costs of backing down, such as providing a rhetorical justification or imposing economic sanctions in lieu of military force (Davies and Johns 2013; Levendusky and Horowitz 2012; Lin-Greenberg 2019; Quek and Johnston 2018; Weiss and Dafoe 2019). These studies have shed light on the agency of signalers in trying to loosen their tied hands *ex post*. However, the agency of opponents remains unexplored.

The opponent's agency is extremely interesting if we view tying hands as a strategic contest between signaler and opponent. If the goal for the signaler tying hands is "to relinquish further initiative, having rigged the incentives so that the other party must choose in one's favor" (Schelling 1960, 37), then the rational response of the opponent should be to return initiative to the signaler, so that it is not trapped in the rigged incentive structure to choose in the signaler's favor. It is in the signaler's interest that its tied hands are tied with a dead knot. Yet, the *ex ante* commitment is not easy to create precisely because the opponent has an incentive to unravel the commitment *ex post*. Schelling (1960, 24–5) illustrates the point with the example of a home buyer tying hands with a third-party contract not to pay beyond a fixed amount and the seller actively seeking out and paying off the third party to release the buyer from the contract. In the context of audience costs, the logic of tying hands requires "statements calculated to arouse a public opinion that permits no concessions to be made," yet "it is by no means easy to establish the commitment … [and] similar activity may be available to the parties on both sides" (Schelling 1960, 28).

If the tied hands of the signaler can be untied, the opponent benefits. Assuming both actors are rational and that creating an absolute commitment is difficult, we can expect a strategic process with the signaler and opponent tying and untying hands. Tying hands will not be a one-shot strategy, but a dynamic process in which the opponent intervenes and influences.

It is common wisdom among international negotiators that opponent actions can affect bargaining space through reverberation on domestic opinion and preferences.[3] Some of the most remarkable reversals in foreign policy have been preceded by opponent actions that reduced potential domestic backlash. Gorbachev's public reassurances helped to open political space for Reagan to reverse his policies and thaw the Cold War (Garthoff 1994). China's gestures and statements in April 1971—publicized through "Ping-Pong diplomacy"—created a positive public image of China in the United States that helped to lubricate Nixon's rapprochement with China against domestic resistance.[4] Seasoned politicians do not force the other side into a dead end with no room to concede, but instead try to calibrate threats and assurances to incentivize the other side to yield. In the 1994 North Korean crisis, for example, the United States tried to ensure its aggressive and conciliatory signals had "the right balance" (Clinton 2004, 591); in the 1996 Taiwan Strait crisis, Clinton wanted the US military deployment to be "without fanfare," as he saw China's leaders trying domestically to "balance between economic pragmatism and aggressive nationalism" (Clinton 2004, 703).

Response Strategies

Consider foreign policy crises where one state publicly threatens to intervene, which constitute the standard setting in most experiments on audience costs (Brutger 2021; Croco, Hanmer, and McDonald 2021; Davies and Johns 2013; Huddleston 2019; Kertzer and Brutger 2016; Levendusky and Horowitz 2012; Levy et al. 2015; Li and Chen 2021; Lin-Greenberg 2019; Nomikos and Sambanis 2019; Quek 2017; Schwartz and Blair 2020; Tomz 2007; Trager and Vavreck 2011). How do opponents in the real world respond to public threats?

First, the opponent may counterthreaten the signaler. When the United States threatened to intervene in Europe during the Second World War, Germany warned the United States to "take note of one thing: every ship, whether with or without convoy which appears before our torpedo tubes is going to be torpedoed … [if] financial hyenas want war … they will get the surprise of their lives" (Hitler 1941). When the United States threatened to sell more weapons to Taiwan, China responded with a threat that it would "take all necessary measures to safeguard its own interests, including imposing sanctions on U.S. companies involved' (*Washington Post* 2019). In the recent crisis in Venezuela, the United States was warned that if it "intends to intervene, they will have a Vietnam worse than what they can imagine" (*New York Times* 2019a).

While a threat for a threat is relatively common, however, the opponent may also choose to reassure instead of threaten. Russia, for example, gave an explicit reassurance to the United States that it would not invade Ukraine during the Crimean crisis (*The Telegraph* 2014). During the Russo-Japanese War, Japan reassured the United States that it had no designs on US possessions in the region (Zabriskie 1973). Besides counterthreatening and reassuring, the opponent may also respond by invoking norms. In the recent Venezuela crisis, Putin warned the United States that "external interference is a gross violation of the fundamental norms of international law" (*New York Times* 2019b). China also regularly invokes the norm of non-interference to fend off US interference in the Diaoyu/Senkaku dispute and in the South China Sea (see, for example, Chinese Ministry of Foreign Affairs 2016; *The Guardian* 2014).

Three response strategies by the opponent are generalized as follows: (1) counterthreat; (2) reassurance; and (3) normative framing.

---

[3]As Putnam (1988, 454) observes: "international pressures 'reverberate' within domestic politics, tipping the domestic balance and thus influencing the international negotiations."

[4]Very soon after Premier Zhou Enlai's meeting with the US players, Washington announced that the twenty-two-year-old trade embargo on China would be terminated, as "Ping-Pong diplomacy had completely changed the political atmosphere" (Chen 2001, 261).

In the first strategy, the opponent *counterthreatens* to impose costs on the signaler. This can create audience costs for the opponent and reduce the audience costs for the signaler under some conditions. When the opponent counterthreatens, it reveals new information about its resolve—or probability of fighting—which changes the expected costs of escalation. In view of the new information, the public may conclude that the issue is not worth further escalation. This argument assumes a rational public that weighs the costs and benefits of escalation, and prefers their leader to de-escalate when escalation is perceived as too costly. If this assumption holds, then audience costs may decrease. Conversely, audience costs may not decrease if the inverse of this assumption holds, that is, if the public is insensitive to the costs of escalation, puts overriding value on the stakes involved, or takes backing down as an indicator of incompetence (Smith 1998) and punishes the leader purely or largely based on perceived incompetence. Which assumption is right in which case is, of course, an empirical question.[5] The literature is split on this point. On the one hand, the public cares about their leader's inconsistency (see, for example, Levy et al. 2015) and incompetence (see, for example, Gelpi and Grieco 2015). On the other hand, the public also cares about the costs and casualties of conflict (see, for example, Gartner 2008; Gelpi, Feaver, and Reifler 2009; Johns and Davies 2019; Mueller 1973), as well as policy substance (see, for example, Chaudoin 2014; Kertzer and Brutger 2016). Some scholars are skeptical that people care more about policy consistency and reputation compared to war and casualties.[6] Others have argued that citizens may understand that it can sometimes be optimal for leaders to gamble and make empty threats in international politics, and then fold their hand rather than go all-in when the cards are not in their favor (see, for example, Desch 2002; Gowa 1999; Ramsay 2004; Slantchev 2006). An important exception is when the stakes involved in the crisis are perceived to be very high, particularly in cases where a country's core interest, such as territorial integrity, is involved. In such cases, however, the credibility of resolve is less likely to be doubted, and the need for signaling through audience costs becomes less relevant as a consequence.[7]

In the second strategy, the opponent *reassures* the signaler that it would not exploit the signaler. The opponent generates "type II" audience costs for itself: the political costs of making an international promise and then backing out (Levy et al. 2015; Quek 2017), which can help to make the reassurance credible. The idea is to persuade the signaler (and its domestic public) to revalue the consequences of backing down—that backing down might not damage the signaler's interests as much as it originally feared. Yet, the reassurance strategy can also work even if the reassurance signal is not fully credible. Even if the public reassurance is ultimately no more than cheap talk, it can still reduce the signaler's audience costs if it helps to "save face" and reduce the loss in national prestige perceived by the public. For example, Snyder and Diesing (1977, 118, 122) highlighted the use of "inexpensive ways of increasing the [loser's payoff] (face-saving) in order to stabilize the situation and prevent the loser from thinking of revenge … perhaps providing some minor or apparent concession … which the loser can use and magnify afterward to obscure the extent of his defeat." Here, the idea is to provide a fig leaf to cover the signaler's backing away from its commitment, and thereby loosen the tied hands of the signaler. It should be

---

[5]Whether the counterthreat strategy is used and whether it works are also empirically testable. The opponent's counterthreat is observable behavior and we do see many counterthreats in the real world. Nevertheless, observational studies of the effect of counterthreats face a similar challenge as observational studies of audience costs: if leaders know that the counterthreat strategy will not work in a particular case, they are less likely to use the strategy, and thus the researcher is less likely to observe the use of the strategy in that case (see Baum 2004; Schultz 2001b; Tomz 2007).

[6]Slantchev (2011, 51), for example, suggests that it would be heroic to assume that people care more about policy consistency compared to "the blood, the destruction of lives and property, and the psychological scars a war invariably inflicts on its participants."

[7]This is a reason why experiments on audience costs almost always use a scenario based on intervention in a foreign country. If the issue at stake is highly salient and important to the signaler (e.g., territorial defense), the credibility of resolve becomes less of an issue. Quek and Johnston (2016) showed that audience costs can diminish in a salient territorial dispute, where the political costs of inaction can overwrite audience costs altogether.

noted that reassurance and counterthreat are not mutually exclusive strategies. If properly timed, the two may work well together. In some cases, a firm counterthreat may be a necessary precondition to offering reassurance.[8]

In the third strategy of *normative framing*, the opponent sensitizes the audience to the signaler's deviance from an existing norm or institution. This strategy opens a new dimension in audience cost theory by intersecting the logic of appropriateness with the logic of consequence. Here, the opponent tries to influence the audience directly through a normative frame instead of indirectly through the audience's observation of the crisis outcome.[9] The costs of backing down may be reduced if the signaler's backing down is perceived to be an appropriate behavior consistent with an existing norm or institution. Crisis outcomes and the valuation of these outcomes can be subject to rhetorical "heresthetics"—the use of language to shape political salience and attract adherents (see also Cohn 1987; Evangelista 2001; Riker 1996). "Backing down" in an international crisis is what leaders and their domestic publics make out of it.[10] By appealing to publicly accepted norms to give meaning to the signaler's escalation or de-escalation, the opponent may delegitimize one action and legitimize the other (see Goddard 2018).[11] If the opponent's normative frame resonates with domestic publics, then escalation will incur the costs of illegitimacy. Norm breaking and anti-institution behaviors are undesirable insofar as the audience values international norms and institutions. Audience costs can be diluted if the public aversion to inconsistency from backing down (based on the logic of consequence in audience cost theory) is counterbalanced by the public aversion to inconsistency from breaking norms (based on the logic of appropriateness in constructivist theory).

In a similar vein, international law and institutions can also offer a mechanism to untie tied hands in a crisis. As Schelling (1966, 120) observes: "[o]ne of the values of laws, conventions, or traditions … is that they provide a graceful way out." Snyder and Diesing (1977, 204) argue that international institutions can reduce the cost of backing down "because conceding to the will of the 'community' is less humiliating than conceding to the will of another state." Institutions like the United Nations (UN) can therefore play a positive role in de-escalating international crises through their influence on domestic and international opinion (Chapman 2007; Grieco et al. 2011; Tago and Ikeda 2013; Voeten 2005). This mechanism formed the basis of President Kennedy's backup plan for de-escalating the Cuban Missile Crisis in 1962. Besides dispatching his brother Robert Kennedy to negotiate with the Soviet Ambassador, President Kennedy also dispatched Secretary of State Dean Rusk to ask Andrew Cordier, a professor at Columbia University who had served as the deputy UN secretary-general, to "suggest to [UN Secretary-General] U Thant that the UN call on the superpowers to withdraw their missiles from Cuba and Turkey…. In case Robert's mission failed, Kennedy believed, this might present him with another way of resolving the crisis peacefully" (Fursenko and Naftali 1997, 281).

The theoretical anchor of tying hands is in the irreversibility of the commitment. The question is whether opponents can actually reduce the costs of backing down and untie the signaler's tied hands through their reactions in the form of counterthreats, reassurances, or normative framing. This is fundamentally an empirical question. I break the question into three hypotheses—one for each strategy—and test them experimentally: Can the opponent's counterthreat (H1), reassurance (H2), or normative framing (H3) reduce the public costs of backing down?

---

[8]If the reassurance is offered too early, the opponent may infer weakness in the signaler's resolve and push the signaler harder (see Snyder and Diesing 1977, 255). I am grateful to an anonymous reviewer for bringing this to my attention.

[9]I thank an anonymous reviewer for highlighting this point.

[10]On the social construction of security concepts, see Wendt (1992), Finnemore (1996), and Katzenstein (1996).

[11]Goddard (2018) argued that a rising power can use a rhetorical frame to legitimate its assertive behavior (e.g., military buildups and economic competition) and render that behavior less threatening. The historical examples include Bismarck's argument that the 1864 invasion of Denmark upheld rather than overturned Concert treaties; and the Monroe administration's argument that U.S. incursions into Florida followed international law on the use of force.

Experiment

To answer this empirical question, I conducted an experiment in January 2020 on a national sample of 1,531 US citizens recruited by Lucid, which used quota sampling to match the Census adult population on gender, age, and income.[12] The experiment applies the same foreign-intervention crisis scenario used in many experiments on audience costs.[13] Respondents saw a scenario in which a country had sent its military to take over a neighboring country and the US president threatened military force if the opponent country continued to invade. The opponent country continued to invade, but the president did not send troops in the end.

Respondents are randomly assigned to the control group or one of the treatment groups. All respondents read the same scenario, except that the treatment scenarios injected an additional sentence. In the counterthreat treatment, the opponent "threatened to make it very costly for the US if the US intervened in the conflict." In the reassurance treatment, the opponent "promised that it would cooperate with the US to ensure regional stability after the war." In the framing treatment based on the non-interference norm, the opponent said that "the conflict is purely between the two countries and that external actors should not be involved in it." We would like to know if the opponent country can use a simple signal—whether by issuing a threat, giving a reassurance, or invoking the non-interference norm—to untie the president's hands and reduce the domestic costs of backing down.

We would also like to know if the actions of *other* external parties—beside the opponent country—can have untying-hands effects. I injected two additional treatments into the experiment. In the UN treatment, the UN Secretary-General said that "it would be inappropriate for the US to unilaterally intervene in the conflict,"[14] whereas in the experts treatment, military experts said that "it would be very costly for the US if the US intervened in the conflict." The former investigates if global institutions like the UN can play a positive role in de-escalating international crises through the mechanism of influencing domestic opinion and loosening tied hands (Chapman 2007; Grieco et al. 2011; Tago and Ikeda 2013; Voeten 2005). The latter examines whether the opinions of other domestic actors can shape the tied-hands signals of a democratic leader (see Schultz 2001a).

The latter treatment has an additional methodological purpose. It is designed to partial out the "provocation effect" in the threat treatment, which can potentially confound our experimental comparison. A threat carries not only a signaling effect that transmits information about the signaler's resolve to impose costs, but potentially also a "provocation effect" that triggers emotions and reactance in the receiver.[15] Parallel assignment of the experts treatment and the threat treatment under a between-subjects design allows us to separate the two potential effects. Both treatments are designed to be symmetric in their *message* (cost of intervention) but not in their *messenger*. The messenger is the foreign opponent in the counterthreat treatment but neutral

---

[12]For validation studies of Lucid samples against National Census benchmarks and US probability samples, see Coppock and McClellan (2019). As the focus is on testing hypotheses rather than making claims about population parameters, previous experiments on domestic audience costs have used non-probability samples recruited through Mechanical Turk or commercial survey companies (e.g., Brutger 2021; Huddleston 2019; Kertzer and Brutger 2016; Levy et al. 2015; Lin-Greenberg 2019; Nomikos and Sambanis 2019; Quek 2017).

[13]Following recent work (Levy et al. 2015; Quek 2017), the experiment abbreviated the scenario by removing the four randomized contextual variables in the original Tomz (2007) vignette: invader's regime type, military strength, motive for conflict, and potential US interests.

[14]See, e.g., BBC (2004).

[15]On the decomposition of the effect of a signal, see Quek (2016, 927). On the causes and consequences of provocation in crisis diplomacy, see Hall (2017) and Cho (2018). Experiments on audience costs using real-world crisis settings suggest that threats may be provocative (Weiss and Dafoe 2019) and potentially counterproductive (Quek and Johnston 2018). On the psychological theory of reactance, see Brehm (1966) and Miron and Brehm (2006). On how reactance explains resistance to threats in crisis diplomacy, see Altman and Powers (2019). Reactance is a motivational state "directed toward the restoration of the threatened or eliminated behavior" when "individuals feel that any of their free behaviors … is eliminated or threatened with elimination" (Miron and Brehm 2006, 10).

(domestic experts) in the experts treatment. If foreign provocation has an effect, we should see different outcomes across the treatments.[16]

The outcome variables are approval rating and perceived reputational impact. Approval rating is measured by asking respondents if they approved or disapproved of the way the US president handled the situation. Reputational assessment is elicited by asking respondents if they believed the United States' reputation has been improved or damaged by the president's handling of the situation. Online Appendix 1 reproduces the full text of the experimental instrument. Online Appendix 2 visualizes the demographic covariate balance across experimental conditions.

Findings

Can the opponent loosen the tied hands of the signaler? Figure 1 tells us that the answer is yes. The figure displays the percentage approval of the leader across the control and treatment conditions, using an experimental comparison of the public costs of backing down similar to past work (Quek and Johnston 2018; Weiss and Dafoe 2019). In the baseline condition, 23 per cent of respondents approved of the president backing down in the crisis. This is comparable to the 22 per cent in Tomz (2007) and the findings from recent experiments that used a simplified Tomz vignette (29 per cent by Levy et al. [2015] and 24 per cent by Quek [2017]).[17]

Now, we turn to the respondents randomly exposed to the counterthreat, reassurance, and normative-frame treatments. Compared to respondents in the control group (23 per cent), respondents in each of the three treatment groups were significantly more approving of the president backing down: 35 per cent in the counterthreat treatment ($p = 0.0026$, $n = 515$), 36 per cent in the reassurance treatment ($p = 0.0011$, $n = 516$), and 41 per cent in the normative-frame treatment ($p < 0.0001$, $n = 512$).[18] The normative-frame treatment has the largest effect in absolute terms, representing a jump of 18 percentage points in public approval for the president.[19] Thus, the opponent has different means to untie hands and make it less costly for the signaler to back down, insofar as the signaler's domestic audience is aware of the words and deeds of the foreign opponent.[20]

We would also like to know if other external actors can untie the hands of the domestic leader. Does an international statement from the UN or a domestic statement from experts reduce the leader's public costs of backing down? Figure 1 shows that, again, the answer is yes. Compared to the control group, in which 23 per cent approved of the president backing down in the crisis, respondents who received the UN treatment and experts treatment cut their president more slack for backing down: 40 per cent approved in the UN treatment ($p < 0.0001$, $n = 512$) and 36 per cent in the experts treatment ($p = 0.0013$, $n = 520$), which translate to increases in public approval for the president by 17 percentage points and 13 percentage points, respectively. Thus, beside the foreign opponent, other external actors—such as domestic experts and international third parties—can also say or do things that contribute toward untying the hands of the domestic leader.

We now turn to how citizens perceive the reputational impact of their president backing down in the crisis. Domestic perceptions of reputational damage matter to US leaders (Brutger and Kertzer 2018), and scholars and policymakers have long been concerned about the reputational impact of backing down (see, for example, Dafoe, Renshon, and Huth 2014; Mercer 1996; Press 2005; Tang 2005; Yarhi-Milo 2018). Figure 2 shows how citizens perceive the impact on the

---

[16]I find no "foreign provocation" effect on public approval (see Figure 1) and reputational assessment (see Figure 2), at least in the foreign-intervention scenario involving an unknown opponent found in standard audience-cost experiments.

[17]See footnote 13.

[18]All tests are two-tailed tests of proportion for percentage differences or two-tailed t-tests for numerical score differences. The approval percentage includes those who said they "lean toward approving," "somewhat approve," or "strongly approve."

[19]As a comparison, the domestic audience costs measured in Tomz (2007, 827) were 12 percentage points in public approval loss (and about 15 percentage points if "leaners" are included).

[20]I discuss the relevance of the domestic media environment in the final section.
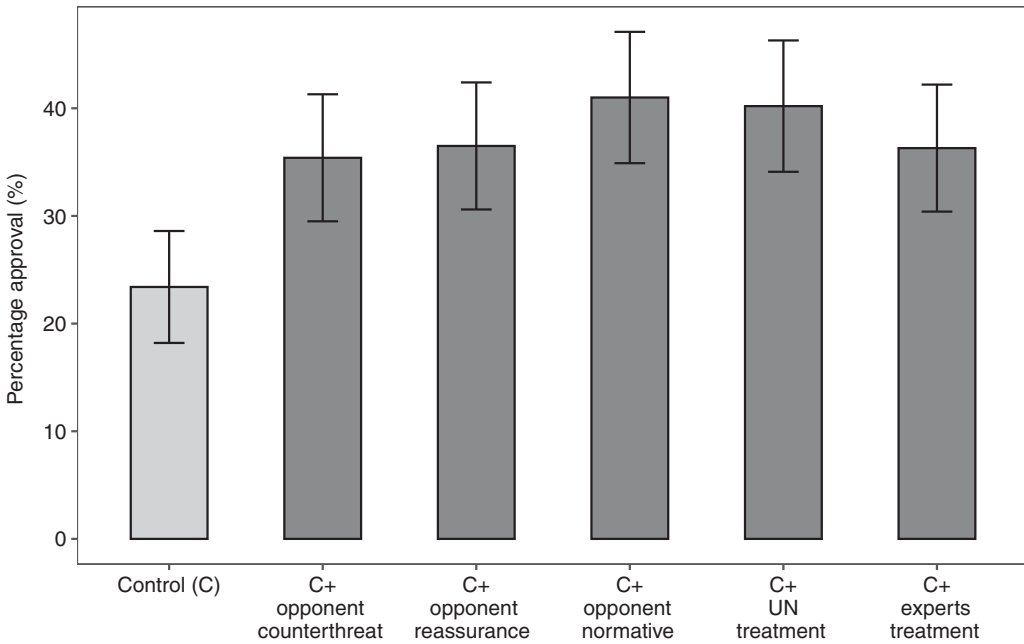
Fig. 1. Percentage approval for the US president (with 95 per cent confidence intervals).

United States' reputation in the control and treatment conditions on a seven-point scale (1 = "damaged a lot"; 7 = "improved a lot"), with 4 at the midpoint ("neither improved nor damaged").[21] Consistent with the results from the approval ratings, reputational impact assessments generally improved when the leader's hands became untied. While the control group assessed the reputational impact of backing down at 2.9, respondents exposed to the foreign opponent's counterthreat and reassurance perceived a more benign impact on their country's reputation at 3.4 ($p = 0.0029$, $n = 515$) and 3.5 ($p = 0.0006$, $n = 516$), respectively. Respondents exposed to the opponent's invocation of the non-interference norm evaluated an impact of 3.2 ($p = 0.0761$, $n = 512$). Meanwhile, those given the UN treatment assessed the reputational impact at 3.5 ($p = 0.0004$, $n = 512$), and those given the experts treatment assessed the impact at 3.3 ($p = 0.0297$, $n = 520$).

In short, not only did the approval ratings improve when the leader's tied hands were loosened, the public's perceptions of the reputational impact also tended to become more benign. It is also interesting to note that respondents did *not* perceive a huge damage to the United States' reputation.[22] Across the different experimental groups, the reputational impact assessments hover in the range of 2.9–3.5, which translate to the United States' reputation being damaged "a little" on the seven-point reputational impact scale.

Finally, I also run regressions on the approval and reputational outcomes as a robustness check. The regression specifications control for different arrays of respondent characteristics, such as age, gender, race, income, education, and party affiliation. Online Appendix 3 shows the ordered-logit estimates of the treatment effects, together with the demographic controls. Across different model specifications, the conclusions remain unchanged (see also Quek 2021b).

---

[21]The measure is related to, but not exactly comparable with, Brutger and Kertzer's (2018) measure of reputational damage. Brutger and Kertzer (2018), which focused on reputational costs, asked "How much damage do you think there would be to America's reputation?" on a five-point scale from "No damage" to "A lot of damage."

[22]At least in this foreign-intervention scenario involving an unknown opponent, which is the standard scenario in audience-cost experiments, including Brutger and Kertzer's (2018) study on reputational costs.
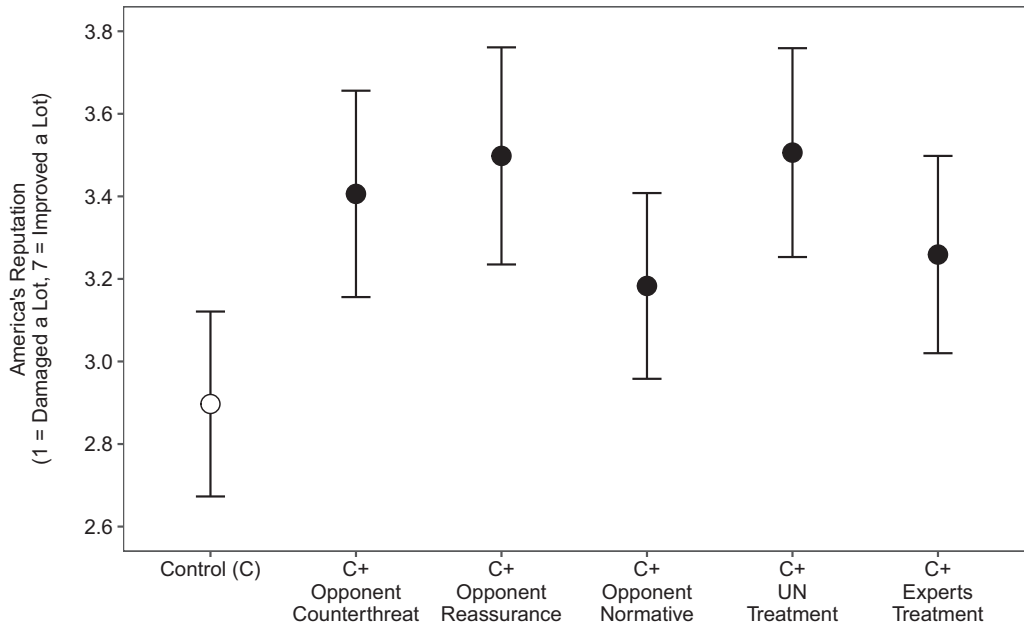
Fig. 2. Perceived impact on the United States' reputation (with 95 per cent confidence intervals).

Implications

To be effective, tied hands must remain tied. I show that opponents can use specific strategies to loosen the tied hands and influence the audience costs of signalers. By counterthreatening, reassuring, and normative framing, the opponent can reduce the domestic costs of backing down for the signaler. This may be one reason why opponents seldom sit still in the face of a public threat. They often respond with counterthreats, reassurances, or normative frames of their own. Such responses are beneficial because loosening tied hands can return to the signaler the freedom to concede. Thus, tying hands is not a one-shot strategy, but a dynamic process in which the opponent has incentive to intervene and influence.

By and large, the literature has assumed tied hands as a fixed commitment and the opponent reacting to the signaler's audience-cost signal as given. I problematized this assumption, showed the signaler's audience costs are sensitive to the actions of the opponent, and identified the inherent difficulties in tying hands. The findings point to a second stage of the research agenda that is potentially very interesting: while "it is worthwhile to consider briefly a model in which practical problems are absent—a world in which absolute commitments are freely available … [t]he most interesting parts of our topic concern *whether and how commitments can be taken*" (Schelling 1960, 26, emphasis added).

Future work may theorize and test the conditions under which specific commitment tactics are robust to the opponent's countermoves. Future research may also delve into the specific mechanisms that support each strategy of untying hands. For example, does the normative framing strategy operate through public aversion to inconsistency (see, for example, Levy et al. 2015), instrumental concerns over reputation (see, for example, Tomz 2007), social preferences for international standing (see, for example, Johnston 2001), or other mechanisms? It will also be interesting to investigate when and why one strategy might work better than another—for the public as a whole and for specific demographics that are politically important.[23]

---

[23]For analyses of treatment effects by specific demographics, see Online Appendix 4.

Another avenue for research is how the signaler responds to its opponent's attempt to untie its hands. Does the signaler counter-respond to retie its hands and reinforce its audience costs? Or, does the signaler take the fig leaf given by the opponent to untie its hands, de-escalate the crisis, and avoid domestic punishment? Under what conditions would we see the opponent untying the signaler's hands, or the signaler *allowing* its opponent to untie its hands? These are interesting questions that require a systematic examination of the empirical record, with a careful eye to differences in regime types and media environments. In an open media environment, while domestic audience costs tend to bite more tightly (see, for example, Choi and James 2007; Potter and Baum 2014; Slantchev 2006; but see also Weeks 2008), the words and deeds of the foreign opponent are also more likely to be transmitted to the domestic audience. All else equal, democracies with media freedom should find it easier to generate domestic audience costs and tie their hands but also harder to insulate themselves from their opponent's attempts to untie their tied hands. Future work may theorize and test these regime–media dyadic combinations and their implications.

Beside extending our general understanding of commitment creation and crisis diplomacy, the findings may also inform specific debates in the literature. For instance, if the opponent's responses—counterthreats, reassurances, or normative frames—can dilute the costs of backing down, this may explain why public threats and audience costs may not work as optimally in the real world as theory suggests in partial equilibrium analysis (Snyder and Borghard 2011; Trachtenberg 2012).[24] If public intervention by the UN can loosen the tied hands of the signaler, then international institutions may play a positive role in de-escalating crises through their influence on domestic and global opinion (see, for example, Chapman 2007; Grieco et al. 2011; Tago and Ikeda 2013; Voeten 2005), despite their hard power constraints.

The findings also have deeper implications for the dynamics of signaling and credibility. Can a tied-hands signal still be informative if the opponent can loosen the signaler's tied hands? My conjecture is yes: the signal is informative if the opponent believes the signaler does not know or believe the tied hands can be untied. Furthermore, even if both sides believe the other may be able to untie hands, the signal is still (partially) informative in correspondence to the perceived probability and estimated effectiveness of the opponent untying hands. We may also presume a process of dynamic interaction: the signaler can respond to the opponent's response with a counter-response to reinforce its audience costs; the opponent can counter-respond to the signaler's counter-response; and so on in an iterative process of tying and untying hands. Given the iterative dynamics, and the incomplete information environment that motivated this process in the first place, we can be sure that everyone can be sure that no one can be sure. Ironically, the informativeness of tying hands derives from uncertainty over the effectiveness of untying hands, when signalers and opponents cannot be sure whether and how much the tied hands can be untied.

---

[24]See also Katagiri and Min's (2019) argument that private statements may be less noisy—and thus potentially more informative—than public statements of resolve.

References

Altman D and Powers K (2019) The Puzzle of Coercion Failure: How Psychology Explains Resistance to Threats. Paper presented at the American Political Science Association Annual Meeting, 29 August to 1 September.

Altman D and Quek K (2021) Do States Signal Resolve by Sinking Costs or Downpaying Costs? Paper presented at the Annual Conference of the International Studies Association, 6–9 April.

Baum MA (2004) Going private: public opinion, presidential rhetoric, and the domestic politics of audience costs in U.S. foreign policy crises. *Journal of Conflict Resolution* **48**(5), 603–631.

BBC (2004) Iraq War Illegal, Says Annan. 16 September.

Brehm JW (1966) *A Theory of Psychological Reactance*. New York, NY: Academic Press.

Brutger R (2021) The power of compromise. *World Politics* **73**(1), 128–166.

Brutger R and Kertzer J (2018) A dispositional theory of reputation costs. *International Organization* **72**(3), 693–724.

Chan ZT, (from Zenobia T. Chan) Liu S, and Quek K (2021) Audience costs and the credibility of commitments. In James P (ed.), *Oxford Bibliographies in International Relations*. New York, NY: Oxford University Press.

Chapman TL (2007) International security institutions, domestic politics, and institutional legitimacy. *Journal of Conflict Resolution* **51**(1), 134–166.

Chaudoin S (2014) Promises or policies? An experimental analysis of international agreements and audience reactions. *International Organization* **68**(1), 235–256.

Chen J (2001) *Mao's China and the Cold War*. Chapel Hill, NC: University of North Carolina Press.

Chinese Ministry of Foreign Affairs (2016) Wang Yi Talks about South China Sea Issue: Facts Will Tell and Justice Will Prevail. 24 May.

Cho H-B (2018) *Tying the Adversary's Hands: Provocation, Crisis Escalation, and Inadvertent War*. PhD dissertation, University of Pennsylvania, USA.

Choi S-W and James P (2007) Media openness, democracy and militarized interstate disputes. *British Journal of Political Science* **37**(1), 23–46.

Clinton B (2004) *My Life*. New York, NY: Alfred A. Knopf.

Cohn C (1987) Sex and death in the rational world of defense intellectuals. *Signs* **12**(4), 687–718.

Conrad CR, Conrad J, and Young JK (2014) Tyrants and terrorism: why some autocrats are terrorized while others are not. *International Studies Quarterly* **58**(3), 539–549.

Coppock A and McClellan OA (2019) Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research and Politics* **6**(1), 1–14.

Croco SE, Hanmer MJ, and McDonald JA (2021) At what cost? Reexamining audience costs in realistic settings. *Journal of Politics* **83**(1), 8–22.

Dafoe A, Renshon J, and Huth P (2014) Reputation and status as motives for war. *Annual Review of Political Science* **17**(1), 371–393.

Davies GAM and Johns R (2013) Audience costs among the British public: the impact of escalation, crisis type, and prime ministerial rhetoric. *International Studies Quarterly* **57**(4), 725–737.

Desch M (2002) Democracy and victory: why regime type hardly matters. *International Security* **27**(2), 5–47.

Evangelista M (2001) Norms, heresthetics, and the end of the Cold War. *Journal of Cold War Studies* **3**(1), 5–35.

Fearon JD (1994) Domestic political audiences and the escalation of international disputes. *American Political Science Review* **88**(3), 577–592.

Fearon JD (1997) Signaling foreign policy interests: tying hands versus sinking costs. *Journal of Conflict Resolution* **41**(1), 68–90.

Finnemore M (1996) Norms, culture, and world politics: insights from sociology's institutionalism. *International Organization* **50**(2), 325–347.

Fischer S (1995) Central-bank independence revisited. *American Economic Review* **85**(2), 201–206.

Fursenko A and Naftali T (1997) *One Hell of a Gamble: Khrushchev, Castro, and Kennedy, 1958-1964*. New York, NY: W. W. Norton.

Garthoff RL (1994) *The Great Transition: American–Soviet Relations and the End of the Cold War*. Washington, DC: Brookings Institution Press.

Gartner SS (2008) The multiple effects of casualties on public support for war: an experimental approach. *American Political Science Review* **102**(1), 95–106.

Gelpi C and Grieco J (2015) Competency costs in foreign affairs: presidential performance in international conflicts and domestic legislative success, 1953–2001. *American Journal of Political Science* **59**(2), 440–456.

Gelpi C, Feaver PD, and Reifler J (2009) *Paying the Human Costs of War: American Public Opinion and Casualties in Military Conflicts*. Princeton, NJ: Princeton University Press.

Gowa J (1999) *Ballots and Bullets: The Elusive Democratic Peace*. Princeton, NJ: Princeton University Press.

Goddard SE (2018) *When Right Makes Might: Rising Powers and World Order*. Ithaca, NY: Cornell University Press.

Grieco J et al. (2011) Let's get a second opinion: international institutions and American public support for war. *International Studies Quarterly* **55**(2), 563–583.

Hall TH (2017) On provocation: outrage, international relations, and the Franco-Prussian War. *Security Studies* **26**(1), 1–29.

Hitler A (1941) Speech before the Reichstag on January 30, 1941. Available from http://www.worldfuturefund.org

Huddleston RJ (2019) Think ahead: cost discounting and external validity in foreign policy survey experiments. *Journal of Experimental Political Science* **6**(2), 108–119.

Johns R and Davies G (2019) Civilian casualties and public support for military action: experimental evidence. *Journal of Conflict Resolution* **63**(1), 251–281.

Johnston AI (2001) Treating international institutions as social environments. *International Studies Quarterly* **45**(4), 487–515.

Katagiri A and Min E (2019) The credibility of public and private signals: a document-based approach. *American Political Science Review* **113**(1), 156–172.

Katzenstein PJ (ed.) (1996) *The Culture of National Security: Norms and Identity in World Politics*. New York, NY: Columbia University Press.

Kertzer J and Brutger R (2016) Decomposing audience costs: bringing the audience back into audience cost theory. *American Journal of Political Science* **60**(1), 234–249.

Kurizaki S (2007) Efficient secrecy: public versus private threats in crisis diplomacy. *American Political Science Review* **101**(3), 543–558.

Leeds BA (1999) Domestic political institutions, credible commitments, and international cooperation. *American Journal of Political Science* **43**(4), 979–1002.

Levendusky M and Horowitz M (2012) Backing down is the right decision: partisanship, new information and audience costs. *Journal of Politics* **74**(2), 323–338.

Levy JS et al. (2015) Backing out or backing in? Commitment and consistency in audience costs theory. *American Journal of Political Science* **59**(4), 988–1001.

Li X and Chen D (2021) Public opinion, international reputation, and audience costs in an authoritarian regime. *Conflict Management and Peace Science* **38**(5), 543–560

Lin-Greenberg E (2019) Backing up, not backing down: mitigating audience costs through policy substitution. *Journal of Peace Research* **56**(4), 559–574.

Mercer J (1996) *Reputation and International Politics*. Ithaca, NY: Cornell University Press.

Miron AM and Brehm JW (2006) Reactance theory—40 years later. *Zeitschrift fur Sozialpsychologie* **37**(1), 9–18.

Mueller JE (1973) *War, Presidents and Public Opinion*. New York, NY: Wiley.

New York Times (2019a) Maduro Sounds Conciliatory but Warns: U.S. Intervention Would Be Worse Than Vietnam. 30 January.

New York Times (2019b) Russia Warns U.S. Not to Intervene in Venezuela as Military Backs Maduro. January 24.

Nomikos WG and Sambanis N (2019) What is the mechanism underlying audience costs? Incompetence, belligerence, and inconsistency. *Journal of Peace Research* **56**(4), 575–588.

Potter PBK and Baum MA (2014) Looking for audience costs in all the wrong places: electoral institutions, media access, and democratic constraint. *Journal of Politics* **76**(1), 167–181.

Press D (2005) *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, NY: Cornell University Press.

Putnam RD (1988) Diplomacy and domestic politics: the logic of two-level games. *International Organization* **42**(3), 427–460.

Quek K (2016) Are costly signals more credible? Evidence of sender–receiver gaps. *Journal of Politics* **78**(3), 925–940.

Quek K (2017) Type II audience costs. *Journal of Politics* **79**(4), 1438–1443.

Quek K (2021a) Four costly signaling mechanisms. *American Political Science Review* **115**(2), 537–549.

Quek K (2021b) "Replication data for: Untying Hands: Deescalation, Reputation, and Dynamic Audience Costs", https://doi.org/10.7910/DVN/8ZFY3U, Harvard Dataverse, V1, UNF:6:n3FTMvVKF2IKECesApINow== [fileUNF].

Quek K and Johnston AI (2016) Crisis Management in the Shadow of Audience Costs. Paper presented at the American Political Science Association Annual Meeting, 1–4 September.

Quek K and Johnston AI (2018) Can China back down? Crisis deescalation in the shadow of popular opposition. *International Security* **42**(3), 7–36.

Ramsay K (2004) Politics at the water's edge: crisis bargaining and electoral competition. *Journal of Conflict Resolution* **48**(4), 459–486.

Riker WH (1996) *The Strategy of Rhetoric: Campaigning for the American Constitution*. New Haven, CT: Yale University Press.

Schelling TC (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schelling TC (1966) *Arms and Influence*. New Haven, CT: Yale University Press.

Schultz KA (2001a) *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

Schultz KA (2001b) Looking for audience costs. *Journal of Conflict Resolution* **45**(1), 32–60.

Schwartz JA and Blair CW (2020) Do women make more credible threats? Gender stereotypes, audience costs, and crisis bargaining. *International Organization* **74**(4), 872–895.

Slantchev B (2006) Politicians, the media, and domestic audience costs. *International Studies Quarterly* **50**(2), 445–477.

Slantchev B (2011) *Military Threats: The Costs of Coercion and the Price of Peace*. New York, NY: Cambridge University Press.

Smith A (1998) International crises and domestic politics. *American Political Science Review* **92**(3), 623–638.

Snyder GH and Diesing P (1977) *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, NJ: Princeton University Press.

Snyder J and Borghard ED (2011) The cost of empty threats: a penny, not a pound. *American Political Science Review* **105**(3), 437–456.

Tago A and Ikeda M (2013) An "A" for effort: experimental evidence on UN Security Council engagement and support for US military action in Japan. *British Journal of Political Science* **45**(2), 391–410.

Tang S (2005) Reputation, cult of reputation, and international conflict. *Security Studies* **14**(1), 34–62.

The Guardian (2014) Obama Says US Will Defend Japan in Island Dispute with China. 24 April.

The Telegraph (2014) Ukraine Crisis: Moscow Assures Pentagon It Will Not Invade. 29 April.

Tomz M (2007) Domestic audience costs in international relations: an experimental approach. *International Organization* **61**(4), 821–840.

Trachtenberg M (2012) Audience costs: an historical analysis. *Security Studies* **21**(1), 3–42.

Trager RF and Vavreck L (2011) The political costs of crisis bargaining: presidential rhetoric and the role of party. *American Journal of Political Science* **55**(3), 526–545.

Uzonyi G, Souva M, and Golder SN (2012) Domestic institutions and credible signals. *International Studies Quarterly* **56**(4), 765–776.

Voeten E (2005) The political origins of the UN Security Council's ability to legitimize the use of force. *International Organization* **59**(3), 527–557.

Washington Post (2019) China Says It Will Sanction US Firms in Taiwan Warplane Sale. 21 August.

Weeks JL (2008) Autocratic audience costs: regime type and signaling resolve. *International Organization* **62**(1), 35–64.

Weiss JC and Dafoe A (2019) Authoritarian audiences, rhetoric, and propaganda in international crises: evidence from China. *International Studies Quarterly* **63**(4), 963–973.

Wendt A (1992) Anarchy is what states make of it: the social construction of power politics. *International Organisation* **46**(2), 391–425.

Wu C (2020) Untying Hands after Public Threats: Signaling Domestic Preferences in International Crisis Bargaining. Working paper.

Yarhi-Milo K (2018) *Who Fights for Reputation in International Politics? Leaders, Resolve, and the Use of Force*. Princeton, NJ: Princeton University Press.

Yoder BK (2019) Retrenchment as a screening mechanism: power shifts, strategic withdrawal, and credible signals. *American Journal of Political Science* **63**(1), 130–145.

Zabriskie EH (1973) *American–Russian Rivalry in the Far East: A Study in Diplomacy and Power Politics 1895–1914*. New York: Praeger.