Win-Win Deescalation

John N. L. Koo
Harvard University, USA

AND

Kai Quek
University of Hong Kong, Hong Kong

Two states are stuck in a conflict because backing down is costly to their leaders. What can they do to exit the conflict peacefully? We propose a novel mechanism that is practically significant but not yet theorized and tested in international relations: *win-win deescalation*. One side "wins" because the other backs down, but the other also "wins" because it successfully frames its action to its public as a win. How effective are win frames? How would the public react to their leader's—and the opponent's—win frames? Using an original experiment in the midst of an ongoing US–China trade war, we identify win frames that can induce the perception of winning and reduce the public costs of deescalation for the Chinese government. The catch, however, is that when America uses the same win frames, it can compromise the Chinese government's effort to deescalate the conflict, even when it is America backing down. Our work sheds light on the role of win frames in international conflict and how, in the eyes of domestic audiences, backing down is not necessarily losing—nor is the other backing down necessarily you winning.

En el caso de que dos Estados se encuentren atrapados en un conflicto porque dar marcha atrás resulta costoso para sus líderes. ¿Qué pueden hacer para salir de este conflicto de forma pacífica? Proponemos un mecanismo novedoso que ya es significativo en la práctica pero que aún no ha sido teorizado ni probado en las relaciones internacionales: la desescalada en que todos ganan. En este caso, un lado ≪gana≫ porque el otro da un paso atrás, pero el otro lado también ≪gana≫ porque logra encuadrar su acción ante su público como una victoria. ¿Cómo de efectivos son los encuadres que rodean estas victorias? ¿Cómo reaccionaría el público a los encuadres que hace su líder con respecto a su victoria?, y ¿a los de su oponente? Utilizamos un experimento original en medio de un conflicto en curso entre Estados Unidos y China, el cual nos permite identificar encuadres de victoria que pueden inducir la percepción de victoria y reducir los costes públicos que tendría la desescalada para el Gobierno chino. Sin embargo, el problema es que cuando Estados Unidos utiliza los mismos encuadres de victoria, puede comprometer el esfuerzo del Gobierno chino para desescalar el conflicto, incluso cuando es Estados Unidos el que da marcha atrás. Nuestro trabajo arroja luz sobre el papel que juegan los encuadres de victoria en los conflictos internacionales y cómo, a los ojos de las audiencias nacionales, el hecho de retroceder no significa, necesariamente, perder, ni el hecho de que el otro de un paso atrás significa siempre ganar.

Deux États sont coincés dans un conflit, car reculer coûterait trop à leur dirigeant. Comment peuvent-ils sortir du conflit de façon pacifique ? Nous proposons un mécanisme inédit, important en pratique, mais pas encore théorisé ni testé en relations internationales : la désescalade gagnant-gagnant. Un camp ≪ gagne ≫ parce que l'autre recule, mais l'autre ≪ gagne ≫ aussi, car il réussit à présenter son action comme une victoire à la population. Quelle est l'efficacité des cadrages gagnants ? Comment la population réagirait-elle aux cadrages gagnants de son dirigeant, mais aussi de l'ennemi ? À l'aide d'une expérience inédite en plein milieu d'un conflit entre Chine et États-Unis, nous identifions des cadrages gagnants qui peuvent donner l'impression de gagner et réduire les coûts publics de la désescalade pour le gouvernement chinois. C'est une situation où le serpent se mord quelque peu la queue néanmoins, car lorsque les États-Unis emploient les mêmes cadres gagnants, ils nuisent aux efforts du gouvernement chinois en vue de désescalader le conflit, même quand les Américains cèdent du terrain. Notre travail met en lumière le rôle des cadres gagnants dans les conflits internationaux et qu'aux yeux du public national, reculer ne signifie pas forcément perdre; un ennemi qui recule ne signifie pas forcément gagner non plus.

Introduction

Two states are in conflict. Both refuse to concede. Each incurs more and more costs as the conflict drags on. But backing down is also costly to the leaders: Not only do they risk damaging their reputation internationally, they may also suffer a backlash domestically. What can the leaders do to get out of the conflict?

This article proposes a novel mechanism of deescalation that is not yet theorized and tested in the literature, but one that is practically significant in international relations (IR). Consider, for example, the Korean War, when Kim Il-sung was confronted with the difficult task of justifying an armistice that signified his failure to "liberate" the whole Korean Peninsula. Instead of admitting failure, Kim publicly

John N. L. Koo is a PhD student and the Raymond Vernon Fellow in the Department of Government at Harvard University. His research focuses on diplomacy and crisis behavior.

Kai Quek is an associate professor at the University of Hong Kong. His research focuses on signaling, conflict deescalation, and the nature and origins of collective beliefs.

Authors' Note: We thank Zenobia Chan, Alastair Iain Johnston, Scott Kastner, Joshua Kertzer, Atsushi Tago, Michael Tomz, Jackie Wong, and session participants at the 2023 MPSA Annual Conference in Chicago, the 2023 APSA Annual Meeting in Los Angeles, and the 2024 IR Reading Group at Harvard for their feedback; Jiaqian Ni and Mengqiao Wang for their assistance; and the University of Hong Kong for its financial support. The project was conceptualized in January 2022 and the experiment was fielded in March 2022. JK proposed the idea of framing victory, framed the hypotheses, and performed the analyses; KQ developed the mechanics of win frames and fielded the experiment.

Koo, John N. L., and Kai Quek. (2024) Win-Win Deescalation. *International Studies Quarterly*, https://doi.org/10.1093/isq/sqae148

declared the armistice agreement a "victory […] against the U.S. imperialist invaders and their stooges" who "are compelled to kneel down before [the North Korean] people and sign the Armistice agreement" (Kim 1981, 410, 440). In the same vein, despite the humiliating withdrawal of nuclear weapons in the Cuban missile crisis, Fidel Castro declared that "Cuba will not lose anything by the removal of the missiles, because she has already gained so much," while Nikita Khrushchev told the Presidium in Moscow that "[w]e saved Cuba" (Fursenko and Naftali 1997, 288, 314). In the US–China trade war, Donald Trump justified calling a truce with China by emphasizing his success in producing "the biggest deal anybody has ever seen" and called it the first "win" for the local steel industry in 25 to 30 years.[1] In each of these cases—and in many others as well—the leader redefined retreat as a win and used it as political leverage. State A "wins" because State B has retreated in the conflict. At the same time, State B also "wins" if the leader can sell its action to its audiences as a "win." This exemplifies *win-win deescalation*.[2]

This mechanism assumes the subjective interpretation of outcomes at the domestic level is not a simple function of the objective facts at the international level. The domestic public's lack of information and knowledge allows media and official statements to play a critical role, both of which may be shaped by elites. The evaluation of international outcomes can also involve a complex analysis of gains and losses—the complexity allows elite framing to play a role in defining the metric of success on which the domestic audience bases its judgment. When the issue is multidimensional, "winning" or "losing" is unlikely to have a unitary definition, opening up the discursive space for leaders to frame their victory. For these reasons, leaders can explicitly or implicitly frame the outcomes of international conflict as their countries "winning." Despite the prevalence of win frames in international politics, however, the literature has not theorized and tested their effectiveness and implications: Do these win frames change the costs of backing down? In what conditions are win frames effective? How would the public react to their leader's win frames—and to foreign win frames?

We test our mechanism of win-win deescalation by designing and fielding an original experiment in China during an ongoing US–China conflict. Our results show that win frames can successfully induce the perception of winning and significantly reduce the domestic costs of deescalation for the Chinese government. Further, the effectiveness of win frames does not vary based on whether they invoke relative or absolute gains, contrary to the predictions of relative gains theory. Finally, when the rival side (the United States) uses the same win frames against China, it imposes significant public approval costs on the Chinese government, even when it is *the United States* backing down. In the eyes of domestic audiences, backing down is not necessarily losing—nor is the other backing down necessarily you winning. Win frames can be a double-edged sword, unless one is able to control how information about the rival's actions and reactions is disseminated to one's audience. The domestic control of information is what authoritarian states like China find easier to achieve than democratic states like the United States.

Our article proceeds as follows: The second section explains our mechanism and propositions, the third our strategy to test our hypotheses, and the fourth our experimental results. We conclude by examining the theoretical and practical implications of our research.

Backing Down, Framing, and Winning

Leaders are deterred from deescalating when backing down is politically costly. Internationally, backing down connotes irresolution (Nalebuff 1991; Carlson 1995; Copeland 1997; Clare and Danilovic 2012; Wolford 2014; McManus 2018; Kertzer 2019; Lupton 2020) and dishonesty (Guisinger and Smith 2002; Jervis 2002; Sartori 2002), compromising the credibility of a leader's future commitments. Domestically, citizens disapprove when leaders impair their country's reputation (Schultz 2001; Tomz 2007) or appear inconsistent (Chaudoin 2014; Levy et al. 2015)—both of which might speak to the leader's incompetence (Smith 1998; Gelpi and Grieco 2015). A large body of experimental research confirms that leaders who commit themselves in a conflict but later back down suffer a cut in their public approval (e.g., Tomz 2007; Trager and Vavreck 2011; Levendusky and Horowitz 2012; Davies and Johns 2013; Kertzer and Brutger 2016). The same logic can also apply to authoritarian states (Weeks 2008; Quek and Johnston 2018; Weiss and Dafoe 2019; Li and Chen 2021).

Yet, tied hands can also be untied (Quek 2022). A recent development in deescalation research suggests that there are at least two broad strategies leaders can adopt to mitigate the costs of backing down. The first is to deploy *rhetorical explanations* for why the country has backed down. Such explanations include those that invoke the consequences of conflict (Quek and Johnston 2018; Weiss and Dafoe 2019) and those that claim new information has necessitated a reassessment of the existing policy (Levendusky and Horowitz 2012; Davies and Johns 2013). The second strategy is to offer *policy alternatives* as a substitute to the original commitment, such as economic sanctions in lieu of military force (Quek and Johnston 2018; Lin-Greenberg 2019; Kohama, Quek, and Tago 2024).

A common thread in this line of research is that post hoc explanations and actions can help the leader soften the stigma of backing down. Backing down in a conflict is a loss to the state's "'honor' and reputation" (Fearon 1994, 581); the domestic audience therefore punishes its leader for backing down. This logic has a familiar ring, given the prominence of audience cost theory in the IR literature (reviewed in Chan, Liu, and Quek 2021). But does backing down really mean that a state has "lost" in the conflict?

We argue this is not necessarily the case—there are at least three aspects of international conflict that provide leaders with domestic leverage despite their decision to back down. First, elites often have privileged access to information (Baum and Potter 2008; Baum and Groeling 2010), which allows them to select information that shines themselves in the best light (Colaresi 2007). Leveraging the information gap, leaders can exaggerate operational successes, downplay the costs their country has suffered, and accentuate the opponent's losses in the conflict. This is most easily achieved in authoritarian nations where media outlets are controlled by the government. China's *People's Daily*, for instance, reported the trade war's negative impact on the United States four times as frequently as it did about the

[1] Trump, Donald J. 2020. "Remarks by President Trump at Signing of the U.S.–China Phase One Trade Agreement." White House Archives. Accessed April 7, 2023. http://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-signing-u-s-china-phase-one-trade-agreement-2.

[2] We use "deescalate" to emphasize decreases in the objective stage of development of the crisis, rather than "concessions" in the political sense. Synonyms such as "concession" and "yielding" carry connotations of subjectively admitting and accepting defeat (see, e.g., Oxford English Dictionary, 2nd ed.: "concession" sense 1a, "yield" senses 5a and b). In our article, we argue otherwise: states can back down from a conflict without incurring the connotations of defeat.

impact on China (Kuang 2023), and in other cases that involve military deployment, casualty figures on the Chinese side are completely omitted.[3] Similarly, in democratic societies, leaders often use their statements and speeches to frame political outcomes. In democracies, however, private media outlets may reduce the informational gap, and opposition parties may contest the rhetoric (Schultz 2001). Real-time media coverage can reduce the government's informational advantage and force it to adapt to the revealed circumstances (Robinson 1999; Gilboa 2005).

Second, the evaluation of international outcomes often involves complex analyses across multiple dimensions. Cues from leaders can help to shape the metrics of success upon which the public bases its judgment (Downs 1957, 223). In many cases, the dispute involves a series of bargains over multiple issues or over multiple dimensions of the issue—and one side is often unable to prevail across all dimensions. Trade-offs and concessions across multiple dimensions render the outcomes of international disputes difficult to assess. Elite cues can guide the public to a particular metric of evaluation. Chinese official rhetoric, for instance, frequently alludes to "strategic gains in the long run" to justify its setbacks and shortfalls in the short run.[4] Conversely, rhetoric about tactical or operational victories may be a useful fig leaf when the outcome is less than ideal (see also Snyder and Diesing 1977, 118–22). In other cases, leaders may invoke relative gains and losses and spotlight the damage that the conflict has inflicted on the opponent.

Third, the definitions and perceptions of winning are often not fixed and unambiguous but flexible and manipulable (e.g., Johnson and Tierney 2006; Mandel 2006; Angstrom 2007). But while the domestic public may not always have a clear idea of what "winning" means, public approval surges when the incumbent's foreign policy is deemed successful, and falls when it is not (e.g., Paterson and Brophy 1986; Bali 2007; Campbell 2010; Hobolt and de Vries 2016). Citizens' perception that their country has "won" or "lost" shapes their political support for the leader.

The Mechanics of Win Frames

Win-win deescalation is a case of deescalation where both sides of the dispute perceive they "win" in the conflict. The concept moves beyond the simple view of "winning" as an absolute *fact*, to a conceptualization of "winning" as a subjective *perception*. The latter is theoretically and practically more significant for explaining human action, as what drives human decisions is not the objective fact per se, but the *perception* thereof. Specifically, what drives public support in win-win deescalation is not winning or losing per se, but the *perception* of winning or losing. Such mutual perceptions of "win" (i.e., "win-win") can occur in a positive-sum or zero-sum setting. In a positive-sum setting, both parties cooperate and give way, and neither is perceived as losing. In a zero-sum setting, only one party backs down. Obviously, for the party backing down, the zero-sum setting is more challenging domestically. We thus focus on win-win deescalation in the zero-sum setting below for two reasons: Theoretically, because it demonstrates how deescalation can be achieved—and the deadlock resolved—unilaterally; empirically, because it is the more challenging case. If win frames work in a hard case where one side backs down but not its opponent, they could be expected to work more generally in less strenuous settings.

We argue that leaders can deploy win frames to make conflict deescalation less costly and hence more likely. Our argument is constructed as follows. First, we develop the concept of win frames by intersecting framing theory with the subjectivity of what winning means. Further, we specify the mechanics of win frames through an adaptation of the conventional expectancy-value model of attitude formation. Finally, we analyze how win frames can induce the perception of winning and reduce the domestic cost of deescalation before summarizing our theoretical predictions into testable hypotheses.

We define *win frames* to mean "schemata of interpretation"[5] that are used to shape the perception of winning and, by extension, political approval. There is a large literature on the political psychology of framing. In Chong and Druckman's (2007) synthesis, the basic premise of framing theory hangs on the fact that an issue "can be viewed from a variety of perspectives and be construed as having implications for multiple values or considerations" (104). Framing is "the process by which people develop a particular conceptualization of an issue or reorient their thinking about an issue"; framing effects are the "changes of opinion" that occur through the "changes in the presentation of an issue or an event" (Chong and Druckman 2007, 104, 106). In the domestic politics of policymaking, framing is the "process by which politicians attempt to raise public support for their policy by encouraging citizens to think about the policies in a particular way" (Tobin, Schneider, and Leblang 2022, 949–50).

Framing is not merely communicating and informing, but a creative and constitutive process in itself (McAdam, Tarrow, and Tilly 2001, 16). In general, frames can produce changes in political preference through two causal pathways (Peterson and Simonovits 2018): first, by affecting the *selection* of considerations that individuals use to evaluate a policy (Druckman and Nelson 2003; Hiscox 2006; Chong and Druckman 2007; Druckman, Peterson, and Slothuus 2013); and second, after a preference is formed, the *salience* of that preference in the evaluation of the political candidate or policy (Chyi and McCombs 2004; de Vreese 2004).

Individuals construct their opinions not from a blank slate but by drawing from the beliefs and memories available in their mind. These beliefs and memories are not equally salient and accessible. At any one point in time, depending on the problem at hand and the context of recall, some become salient and accessible but not others. Framing can therefore operate at three levels: "by making new beliefs available about an issue, making certain available beliefs accessible, or making beliefs applicable or 'strong' in people's evaluations" (Chong and Druckman 2007, 111).

However, framing cannot be operative everywhere and on everything. The elasticity of reality varies. We assume that framing is operative when there is *multidimensionality* in the issue or concept and *contestability* in or between dimensions. Dimensions refer to the characteristics or properties that constitute the issue. If an issue comprises multiple dimensions—even if each dimension is in itself incontestable—how the different dimensions are assessed relative to and *between* themselves can be contestable. Framing is inoperative when an issue is unidimensional and that dimension is incontestable. If we are evaluating physical weight and physical weight only, framing to change every-

[3] See, for instance, the *People's Daily*'s summary of the 1979 Sino-Vietnam War on March 17, 1979, available at <cn.govopendata.com/renminribao/1979/3/17/1>

[4] "中美在磋商中开辟解决问题的新路径 [China and the U.S. Breaks a New Path in Negotiations]." China Radio International, February 1, 2019.

[5] (Goffman 1986 [1974]) conceptualized frames as frameworks or "schemata of interpretation" (21).

one's mind that one pound is heavier than half a pound is difficult, if not impossible. Joe Biden and Donald Trump can frame themselves as "good" leaders because what constitutes a "good" leader is multidimensional and contestable. Both leaders, however, will find it difficult, if not impossible, to frame themselves as 30 years old or 7 feet tall, insofar as we see physical age and height as unidimensional and incontestable (if we find this framing implausible and ridiculous, that is precisely the point). Similarly, the French government will find it difficult to frame itself as winning the 2022 FIFA World Cup when it is unambiguous that—given the rules of the game—Argentina had won. The French, however, could frame themselves as the winner on "moral grounds" or in the "strategic sense," insofar as the "moral" or the "strategic" are multidimensional and contestable.

The efficacy of win frames, therefore, rests on the multidimensionality and contestability of what "winning" means. That "winning" is multidimensional and contestable in or between its dimensions is well supported by research on the meaning of war victory. As Mandel (2006, 4) noted, "[c]onflicting understandings abound of victory"; and indeed, "all students of history must be struck by the ambivalence, irony, or transience of most military victories" (Bond 1996, 1). Angstrom (2007) analyzed the ambivalent views that constitute the US government's understanding of victory in its war against terrorism, involving military definitions and indicators such as casualties as well as political goals and outcomes such as regime change. "Victory" can be defined in many ways and in different dimensions, from the military, economic, and diplomatic to the political, social, and informational (Mandel 2007, 20–30), and "whether the focus is on a fixed end state or a fluid cost-benefit ratio, complicating matters is the seemingly inescapable subjectivity of victory, wherein objective criteria … seem to be playing less of a role [and instead] perceptual bias and manipulation appear to dominate interpretation of military triumphs" (Mandel 2006, 9).

In particular, Johnson and Tierney (2006) observed that the public seldom has a consistent standard of victory, leading to divergent perceptions of victory across different military operations: The American public largely considered the *Mayaguez* offensive a success, even though more US soldiers were killed than the hostages rescued; the Battle of Mogadishu in 1993 was generally thought of as a humiliating loss, even though the United States successfully rescued the survivors and inflicted more casualties on Somalian rebels than what the United States suffered.[6] Perceptions of victory are influenced not only by "score-keeping" based on a listing of material gains and losses but also by what Johnson and Tierney (2006, 38) called "match-fixing" through "three types of phenomena rooted in human nature": "mindsets; salient events; and social pressures." In a similar vein, Mandel (2006) explained the distorted perceptions of military victory through "[t]hree principal bodies of psychological theory": "(1) selective attention—ignoring incoming information that contradicts preexisting images; (2) wishful thinking—focusing just on positive outcomes where desires take precedence over expectations; and (3) cognitive bolstering—seeking out further evidence to enhance the credibility of preexisting beliefs" (75). While the specific psychological explanations are diverse and varied, all of them expand on the same general theme: "Victory is in the eye of the beholder" (Johnson and Tierney 2007, 46).

After establishing the theoretical plausibility of win frames, we now turn to how they work. The conventional expectancy-value model of attitude formation can provide a model of the framing process (Ajzen and Fishbein 1980; Nelson, Oxley, and Clawson 1997), as described in Chong and Druckman (2007). The model posits that attitude toward an object can be quantified as $\sum v_i w_i$, where v_i is the evaluation of the object on attribute i, and w_i is the salience weight ($\sum w_i = 1$) of attribute i. The attribute i can describe a dimension, a consideration, a value, or a belief (Ajzen and Fishbein 1980; Riker 1990; Zaller 1992; Sniderman 1993).

Adapting this framework and applying it to win frames, we conceptualize one's overall evaluation of a foreign policy outcome (e.g., retreat or concession) as the result of a combination of positive and negative evaluations v_i of the outcome on different dimensions i. Suppose for simplicity that two dimensions, political ($i = 1$) and economic ($i = 2$), are relevant, and a political concession (thus $v_1 < 0$) is made for a positive economic purpose (thus $v_2 > 0$). The overall evaluation of the outcome depends on both the relative magnitudes of v_1 and v_2, and the relative salience weights (w_1 and w_2) assigned to each dimension. In this model, win frames can operate on attitudes through different potential pathways: by introducing a new dimension i (e.g., the moral dimension, $i = 3$), modifying the evaluation of an existing dimension i (e.g., the magnitudes of v_1 and v_2), or changing their relative salience weights (e.g., w_1 and w_2).[7]

In this article, the agents deploying the win frames for deescalation are political leaders, and an individual's susceptibility to these win frames is also influenced by the individual's relationship with the leader. Individual citizens tend to follow the cue of elites they trust, especially in issue domains where they lack the relevant knowledge or information (Sniderman, Brody, and Tetlock 1991; Zaller 1992; Lupia 1994; Berinsky 2007). Individual evaluations of foreign policy are influenced by events and their material calculations (e.g., Mueller 1973; Gartner and Segura 1998; Burk 1999; Feaver and Gelpi 2011), but the influence operates in the context of the positions and frames of the politicians whom citizens take as their reference point (Zaller 1992; Berinsky 2007). In some political contexts, such as a liberal democracy like the United States—though not in others, such as an authoritarian state like China—individuals may be exposed to competing frames from competing politicians (Sniderman and Theriault 2004).

How can leaders use win frames to make it easier for them to deescalate a conflict? We propose that leaders are more likely to back down in a conflict when they believe that they can use win frames to create the domestic perception of winning *despite* backing down internationally. Rather than admitting defeat or accepting the necessity of retreat at face value, leaders frame their action as a win or—at least—not a loss for their country. A win frame is effective if it can make more salient and accessible the dimensions of the issue that citizens use to formulate their opinion, which are consistent with the leader's definition of winning and which can activate the citizens' positive evaluation and approval. In turn, the leader becomes more likely to retreat in the conflict because they believe they are less likely to suffer a domestic backlash for backing down. We visualize the theoretical mechanism in figure 1.

[6] For an account on elite perception of military victory, see Mandel (2006), Martel (2007), and Bartholomes (2010).

[7] Chong and Druckman (2007, 105) explained that while the value expectancy model is "an idealized conception of an attitude as a summary of a definable set of beliefs that an individual holds about a subject," its "general assumption that an individual can place different emphases on various considerations about a subject is a useful abstraction for discussing the psychology of framing."

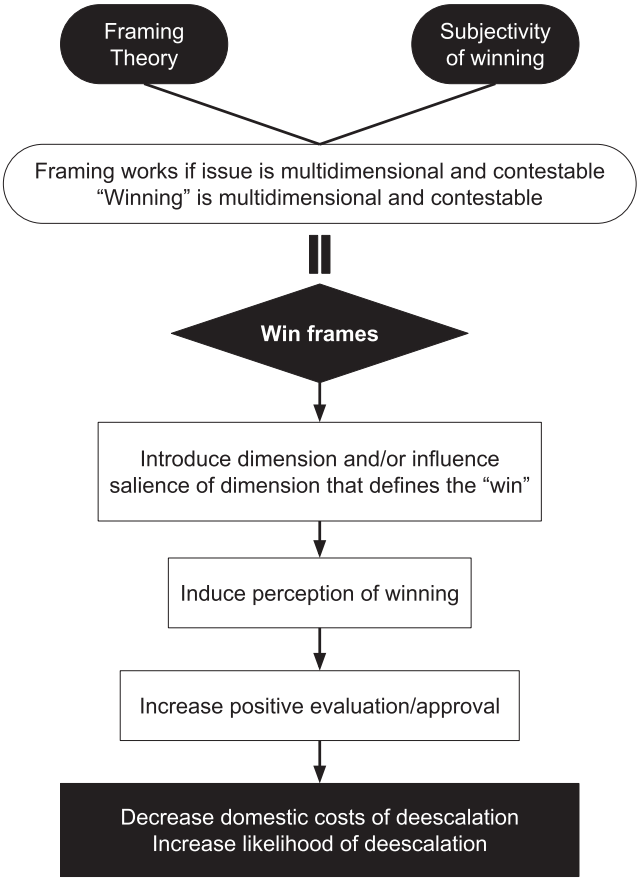

Figure 1. Theoretical mechanism

Win frames induce the perception of winning and cut the costs of backing down by introducing a new dimension and/or influencing the salience of the dimension that defines the "win." It informs the selection and salience of the yardstick that measures the outcome. Of course, many dimensions or standards exist with which one can evaluate the outcome of a conflict. At the general level, however, most of them are rooted in the basic questions of whether the state has made an *absolute* gain, whether it has gained *relative* to the opponent, or whether the objectives of the government have been achieved.

Here, politicians are expected to select and/or make salient a dimension or standard by which the outcome appears favorable. Joe Biden, for example, claimed the hasty retreat of 120,000 American troops from Afghanistan a success by framing it as an operational victory—the "biggest airlift in history [… that] no nation has ever done anything like it in all of history"[8]—but eliding the chaos the retreat brought and how it indirectly allowed the Taliban to retake power. In addition, when leaders claim a win, they are also implying that their true objectives are fulfilled. The "true objectives" often involve multiple dimensions that are ambiguous and contestable. Mandel (2006, 5) suggested, for example, that "[v]irtually all governments and military establishments initiate war with some sort of predetermined objective end state identified, but frequently this is ambiguously stated at the outset." The objective truth value behind a subjective claim of winning is beside the point; indeed, its lack thereof can theoretically contribute to the efficacy of framing. By framing a win and its consistency with the government's "true objectives," the win frames help leaders to deemphasize the inconsistency associated with backing down (Fearon 1994; Tomz 2007; Levy et al. 2015; Kertzer and Brutger 2016; Quek 2017; Takei and Paolino 2023) and control the costs of backing down. While Chong and Druckman (2007) found it "troubling" when policies are "built around frames only because they are known to resonate with the public and not because they address central features of the issue" (111), in the context of conflict deescalation, win frames may enable leaders to circumvent the very features of the issue that are obstacles to peace.

> **Hypothesis H1:** All else constant, win frames should increase public support for how the government has handled a conflict relative to the counterfactual where win frames are absent.

Are some win frames more effective than others? The IR literature suggests that states construe material gains in one of two ways: absolute or relative. Correspondingly, win frames can invoke absolute gains or relative gains.

Previous studies have presented divergent views on the importance of absolute and relative gains along the theoretical cleavage between realist and neoliberal institutionalist theories. On the one hand: "Realists argue that the general insecurity of international anarchy leads states to worry not simply about how well they fare themselves (absolute gains) but about how well they fare compared to other states (relative gains)" (Snidal 1991, 703; see also Grieco 1988), and because the relative power and influence of a country depend on the latter rather than the former, "relative gain is more important than absolute gain" (Waltz 1959, 198). On the other hand: "Neoliberal institutionalism assumes that states focus primarily on their individual absolute gains and are indifferent to the gains of others" and that "this focus on absolute gains is usually taken to mean that a state's utility is solely a function of its absolute gain" (Powell 1991, 1303).

Reality is likely to reside between the two theoretical ends. The importance of relative gains is not a constant but a variable (Powell 1991; Snidal 1991; Rousseau 2002; Yeung and Quek 2022). In fact, while Waltz (1979) argued that in a self-help system "competing parties consider relative gains more important than absolute ones," he also proposed that "[a]bsolute gains become more important as competition lessens" (195). In the context of great-power competition, because "one state may use its disproportionate gain to implement a policy intended to damage or destroy the other," the concern is "not 'Will both of us gain?' but 'Who will gain more?'" (Waltz 1979, 105). Following this line of reasoning:

> **Hypothesis H2:** All else constant, in the context of rivalry, win frames based on relative gains should have greater support than win frames based on absolute gains.

How would citizens react to win frames by *foreign* states? Here, the logic is straightforward: If the two states are rivals, then any declaration of "winning" by one would lead to an uncomfortable inference of "losing" for the other. If so, a win frame by one side is equivalent to a loss frame imposed on the rival side, leading to reverse effects.

> **Hypothesis H3**: All else constant, citizens exposed to a win frame from the opponent country should be less supportive of how their government has handled the conflict compared to citizens who are not.

[8]Biden, Joe. 2021. "Remarks by President Biden on the End of the War in Afghanistan." White House. Accessed October 8, 2024. https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/08/31/remarks-by-president-biden-on-the-end-of-the-war-in-afghanistan.

In the real world, things may not be as neat as they seem, and there could be unintended consequences for the opponent. Hypothesis H3 sets up a "stress-test" of our mechanism to establish its limits and investigate side effects. It helps us to understand holistically the "system effects" (Jervis 1997) instead of restricting ourselves to only one slice of the system and looking at only one direction of the interaction, as in Hypotheses H1 and H2.

Research Design

We designed an experiment in China to evaluate our hypotheses. The experiment was embedded in a survey that was fielded over the Internet in March 2022, with a quota sample of Chinese citizens to match the 2020 National Census on sex, region, age, and ethnicity. Respondents were randomly assigned into one of twelve groups. Details on our sample demographics and the covariate balance across experimental groups are reported in online appendices A and B, respectively.

We focus on China for three reasons. First, the US–China dyad is politically highly consequential. A direct confrontation between China and the United States, the two strongest economic and military powers in the world, will have deep spillover effects on international security and global supply chains, affecting almost everyone. Understanding how tensions between the two superpowers can be eased is therefore imperative. Second, China presents a hard test for our mechanism, insofar as the Chinese public is nationalistic. The historical trauma of the "Century of Humiliation" continues to evoke strong emotions among the Chinese public (Callahan 2004; Wang 2008).[9] If a mechanism helps to placate a nationalistic public, who is expected to be more averse to backing down in the face of a foreign foe, one may expect the mechanism to also be efficacious with other publics that are less fervently nationalistic. Finally, studies have shown that the Chinese government is highly sensitive to how the public feels (e.g., King Pan, and Roberts 2013; Stockmann 2013; Roberts 2018; Gries and Wang 2020; Fang, Li, and Liu 2022; Potter and Wang 2023). While China is not a democracy, public sentiments can still be influential. Perceived grievances at the international level have historically been triggers for public protests. These protests are a security risk to the government, as they can lead to a cascade of anti-regime protests across the nation, increase the costs of repression, and create fissures among party elites (Kuran 1998; Weiss 2014). Indeed, perhaps no other country has invested as many resources as China to monitor and analyze public opinion.[10] Recent experimental work that surveyed the Chinese public to understand the domestic dynamics of China's foreign policy behavior includes Bell and Quek (2018), Weiss and Dafoe (2019), Fang and Li (2020), Li and Chen (2021), Dafoe et al. (2022), Ko (2022), Aksoy, Enamorado, and Yang (2024), Kertzer, Brutger, and Quek (2024), and Myrick and Wang (2024).

We designed our experiment to test win-win deescalation in the context of the US–China trade war. The US–China trade war has been extremely consequential to the world, and thus understanding the domestic dynamics of deescalation in this context will be valuable to both policymakers and scholars alike. While the US–China trade war does not translate directly into battlefield violence, it is a high-stakes conflict that involves costs in the magnitude of trillions[11] and affects countless families and livelihoods that depend directly or indirectly on US–China economic relations. The implications of backing down differ at different rungs of escalation: Backing down should be less costly at lower stages compared to higher stages of escalation, where public confrontations and domestic attention generate greater costs for backing down.[12] While the US–China trade war is not a militarized dispute, it is a high-stakes conflict that has escalated rung by rung—across months of news headlines—in the tit-for-tats between the United States and Chinese governments. The public confrontations and domestic attention over the US–China trade war thus create potentially higher costs for backing down.[13] Our experiment is thus designed to test if win-win deescalation can operate in a high-stakes and publicly escalated conflict.

Our experiment began with an introductory screen that set the scene in the year 2025, where China is engaged in a trade war with the United States. We set the scenario in the future to avoid introducing confounds from the politics of the day.[14] Respondents learned that substantial tariffs have been imposed by each side on the other's economy, and the tariffs have taken a toll on both economies. Thereafter, the respondents were randomly assigned to read one of two mirror scenarios, Scenarios A and B.

Scenario A, administered to Experimental Groups 1–6, tests how Chinese citizens react to win frames by their own government. Respondents read that China has backed down in the conflict and declared the end of the trade war.[15] Outside of the control group (Group 1), Groups 2–6 randomly received one of five win frames, in which China declared that it has won because it: (a) "gained more than the U.S. in the conflict" (*Relative Gain*); (b) "suffered less losses than the U.S. in the conflict" (*Relative Loss*); (c)"gained more than if it had not fought the trade war" (*Absolute Gain*); (d) "has strengthened its international position with the trade war" (*Strategic Gain*); or (e) "has weakened the U.S.'s international position with the trade war" (*Strategic Damage*). Following the theoretical premises that we developed earlier, we should expect public support for how the government has handled the conflict to be higher among respondents exposed to these win frames than those who were not (Hypothesis H1). If relative gains theory is correct in that relative gains are more salient than absolute gains in the context of great-power competition, we should expect the win frames based on relative gains or losses to prevail over the win frame based on absolute gains (Hypothesis H2).

Scenario B (Groups 7–12) mirrors Scenario A and tests how Chinese respondents react to *foreign* win frames. Here, the United States backs down with no win frame (Group 7)

[9]Note, however, that subnational variations exist in the level of patriotism and willingness to mobilize on grievances in China. See Duan (2017), Ong and Han (2019), and Wallace and Weiss (2015).

[10]Weiss and Dafoe (2019, 964) noted that "the government employs more than two million analysts to monitor internet sentiment and win the 'guerrilla battle' in the 'mass microphone era,' according to the head of the People's Daily Public Opinion Monitoring Unit."

[11]For example, Amiti, Kong, and Weinstein (2020) estimated that US and Chinese tariff announcements alone would have wiped out US$1.7 trillion of firm value in a partial sample of US-listed companies.

[12]We are grateful to our anonymous reviewer for a discussion on this point.

[13]Indeed, the US–China trade war may be a context that is more immediate and salient to citizens and their livelihoods than the hypothetical US foreign military intervention scenario in typical audience-cost experiments.

[14]Setting the scene in 2025 would allow respondents to think beyond the tenure of the current Biden administration and beyond the specific time they were in (March 2022).

[15]We left open what the government has done specifically to back down. For our experimental test to operate, what is important is that people know explicitly their government has backed down, which is the focus of our experiment. Our open responses reveal that most respondents understood this to mean giving way or ceasing to insist on certain demands. The investigation of different policies and degrees of backing down will be an interesting avenue for future research.

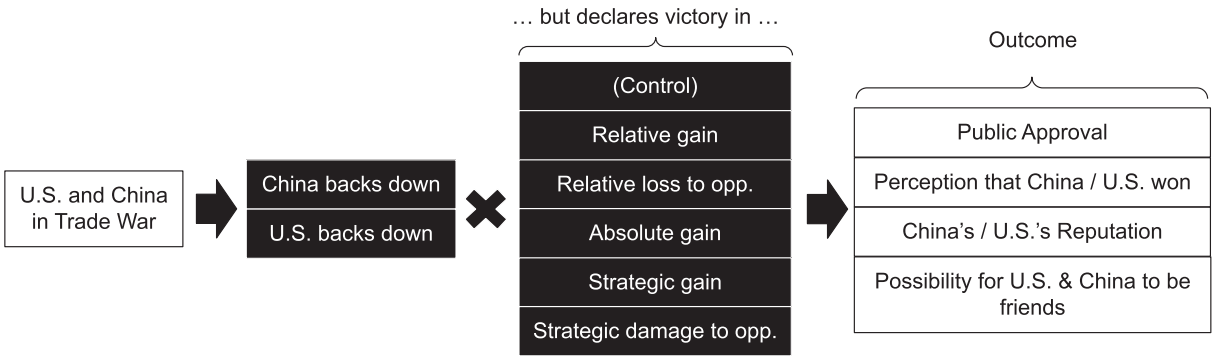

Figure 2. Flow of experiment

or with one of the five win frames as stated above (Groups 8–12). Scenario B matches the language in Scenario A, except that it switches "China" with "the United States." If Hypothesis H3 is correct, public approval should decrease in all treatment groups that received a foreign win frame, with an especially pronounced decrease from win frames based on relative losses and strategic losses, where the losses on China are emphasized by the foreign government.

Our main dependent variable is approval of how the government has handled the conflict. Given the political environment in China and the need to mitigate political sensitivity biases, we avoided asking our Chinese respondents whether they support or oppose the government. Rather, we asked whether the respondent agrees or disagrees that "China handled the incident very well" on a 7-point scale from 1 ("strongly disagree") to 7 ("strongly agree"). Thereafter, respondents can opt to explain why they approved or disapproved in an open-response question.[16]

To check whether the frames were indeed effective in inducing the perception of winning, we also asked respondents if they agree that China or the United States has won, on a seven-point scale from 1 ("strongly disagree") to 7 ("strongly agree"). We also included two secondary questions about reputational changes[17] and perceptions on the future of US–China relations to explore the wider implications of the win frames. Online appendix C reproduces the experimental text and its translation in English. The flow of the experiment is summarized in figure 2.

Findings

We begin by analyzing the baseline effect of win frames. For ease of interpretation, we code approval as the proportion of respondents who strongly agree, agree, or lean towards agreeing that their country "has handled the incident very well" in each experimental group (see online appendix C). Replicating the analyses with the original ordinal scores yields the same conclusions.

Are win frames effective in diluting the domestic costs of backing down (Hypothesis H1)? Figure 3 shows that they are. All five win frames by the Chinese government produced significant increases in public approval compared to the control, where no win frame was given. The effect magnitude was largest in the relative losses frame (where the United States was said to lose more than China in the conflict), which increased public approval from 64.5 percent[18] to 83.9 percent (diff. = 19.4 percentage points, $p < 0.001$, $n = 546$);[19] and smallest in the strategic losses frame (where the US international position was weakened), with approval increased by 7.7 percentage points ($p = 0.051$, $n = 557$). These results support our theoretical proposition that domestic win frames are politically effective (Hypothesis H1).

Are some win frames more effective than others? We find no statistically significant difference in approval across the five win-frame groups, which contradicts the hypothesis (H2) that win frames based on relative gains are more salient than those based on absolute gains.

We conducted additional analyses to check if increases in perceptions of winning were indeed driving the treatment effects. If our mechanism about win frames is correct, we should observe the perception that China has won to be correlated with a win-frame treatment and such a perception to be correlated with the outcome (approval).

Table 1 shows this is indeed the case. From logit regressions of approval against a treatment dummy (1 if any win frame is received and 0 if otherwise) and the respondent's perception that China has won (ordinal scale from 1 to 7), the effects of both the treatment dummy and perception are statistically significant and robust to socio-demographic differences. The effect sizes for both the treatment and perception of winning are also large.

[16]Our quantity of interest is the respondent's evaluation of the government's overall performance in the whole episode. But since the metric by which a respondent evaluates the government is multidimensional (which is partly what enables win frames to be operative in the first place), closed-response data may not be able to tell us which aspect contributed more or less to the respondent's overall assessment. We thus include open-response questions to give us insight into the underlying heterogeneity—the analysis of which is presented in the next section.

[17]Our reputational measure was conceptualized and designed following previous experimental work by Brutger and Kertzer (2018) and Quek (2022), where the focus is on how citizens perceive the reputational impact of backing down in an international conflict. The measure taps on the concept of reputation cost that is part of audience cost theory, where backing down is costly because "it gives domestic political opponents an opportunity to deplore the international loss of credibility, face, or honor" (Fearon 1994, 581). Here, the concept of reputation does not necessarily refer to specific reputations of resolve or honesty, but the generalized concept of reputation as a valued attribute attached to the state's "credibility, face, or honor" *as perceived by the public*, and which can tap into different logics and types of reputations, as Brutger and Kertzer (2018) have shown.

[18]Li and Chen's (2021) audience-cost experiment in China with a humanitarian intervention scenario—which is more comparable to ours than experiments with military crisis scenarios—has a comparable baseline approval of 58.2 percent.

[19]All *t*-tests reported in this section are two-tailed. All differences in means are significant on the seven-point ordinal scores: relative gains (diff in means = 0.48, $p < 0.001$, $n = 556$), relative losses (diff in means = 0.70, $p < 0.001$, $n = 546$), absolute gains (diff in means = 0.54, $p < 0.001$, $n = 560$), and strategic gains (diff in means = 0.62, $p < 0.001$, $n = 546$), and strategic losses (diff in means = 0.45, $p = 0.001$, $n = 557$). The results hold when proportions of respondents who approved are used as the dependent variable instead—except the strategic losses frame at $p = 0.051$.

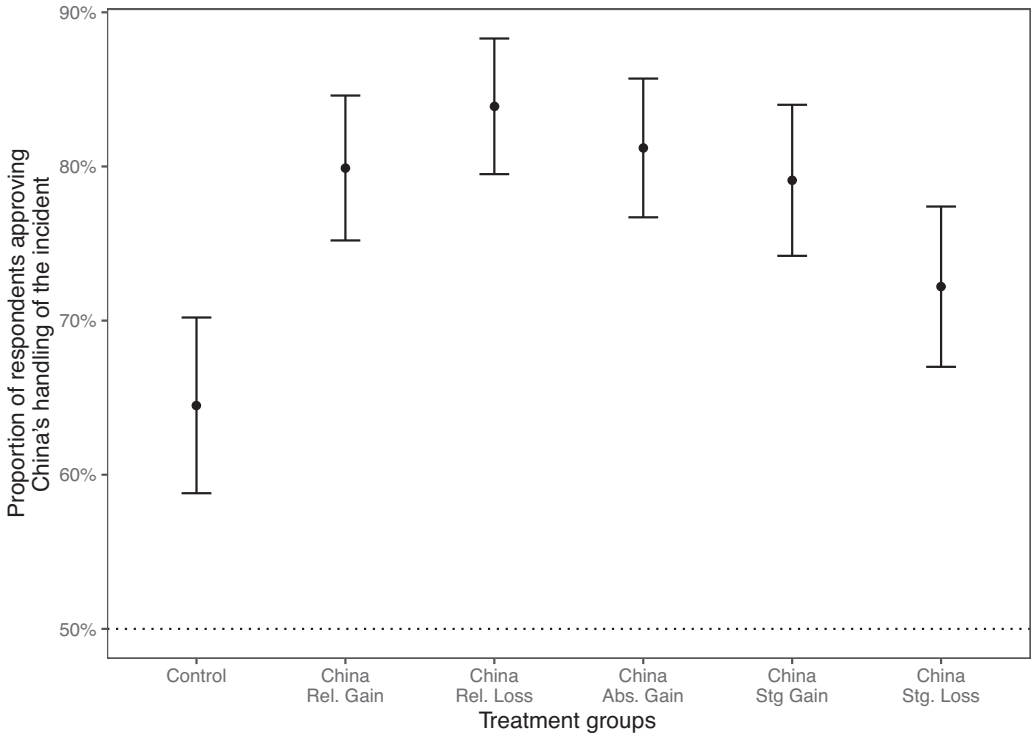

Figure 3. Approval for China's handling of the conflict (China backs down)

Table 1. Logit regressions for public approval, treatments, and perceptions of winning

	Dependent variable				
	Binary approval ({0,1})				
	(1)	(2)	(3)	(4)	(5)
China wins (treatment)	0.742***		0.681***		0.573***
	(0.143)		(0.148)		(0.180)
Perception of winning		0.908***		0.967***	0.961***
		(0.051)		(0.056)	(0.056)
Education			−0.028	0.043	0.049
			(0.058)	(0.066)	(0.066)
Hawkishness			0.042	0.036	0.041
			(0.050)	(0.061)	(0.062)
Income			−0.000*	0.000*	0.000*
			(0.000)	(0.000)	(0.000)
Sex (1 = male)			−0.284**	−0.252*	−0.266*
			(0.122)	(0.145)	(0.146)
Constant	0.596***	−3.146***	0.851**	−3.647***	−4.118***
	(0.126)	(0.243)	(0.366)	(0.481)	(0.510)
Observations	1,673	1,671	1,594	1,592	1,592
Log likelihood	−893.179	−675.360	−846.743	−626.496	−621.535
Akaike Inf. Crit.	1,790.358	1,354.721	1,705.486	1,264.991	1,257.071

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In sum, our experiment suggests that even concise claims about winning—without data or justification—can be effective in eliciting perceptions of victory and increasing public approval despite their government backing down.[20] The general effectiveness of win frames might have been lubricated by motivated reasoning, insofar as citizens tend to prefer seeing (perceiving) their own country winning rather than losing. It would be interesting for future work to investigate the specific metrics, if any, that different individuals have in mind when they evaluate win frames.

However, while win frames may be effective in increasing approval from one's own public, they may generate side effects on the opponent's public. To check whether win frames by a *foreign* rival will hamper one's ability to deescalate, our experiment had randomly allocated, under a between-subjects design, half the sample to read Scenario

[20]We thank an anonymous reviewer for this suggestion.

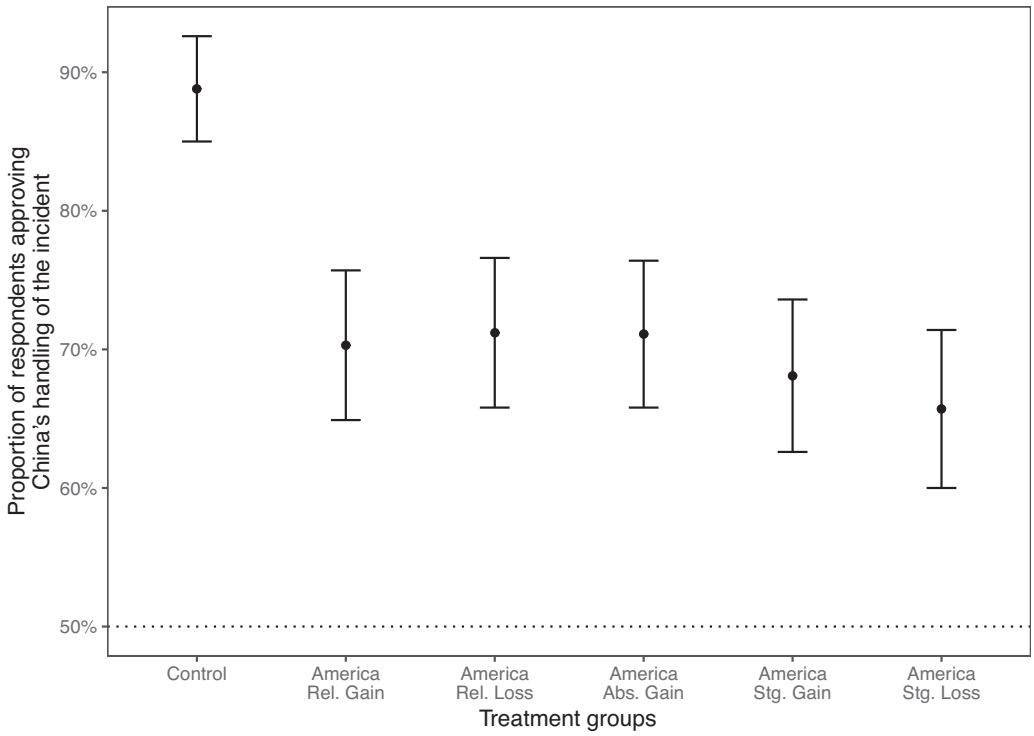

Figure 4. Approval for China's handling of the conflict (US backs down)

B, where Chinese respondents reacted to win frames from the United States. This allows us to test the hypothesis (H3) that foreign win frames would decrease public approval (see the section "Backing Down, Framing, and Winning"). Our results support the hypothesis, as figure 4 shows. All American win frames led to decreases in Chinese public approval by around 20 percentage points. The effect size was largest in the strategic losses frame (diff = –23.2 percentage points, $p < 0.001$, $n = 543$) and smallest in the absolute gains frame (diff = –17.7 percentage points, $p < 0.001$, $n = 553$).[21] Compared to Scenario A, the United States backing down and not offering a win frame (approval at about 88.8 percent) is a more palatable outcome to Chinese citizens than China backing down and offering a win frame (maximum approval at 83.9 percent). But if the United States deploys a win frame (minimum approval at 65.7 percent), the Chinese government may be worse off in approval than if it had backed down and offered a win frame instead.[22]

This suggests that one's win frame is not a magic bullet—it can also affect the opponent, under conditions where foreign win frames are successfully and saliently publicized in the opponent's domestic media. There is thus a potential side effect that generates "externalities" on the opponent—but it occurs only under rather stringent conditions that may not always obtain, given that domestic media could be tightly controlled (in authoritarian regimes) or flooded with multiple messages and messengers contending for public attention (in democratic regimes).

Probing further, we asked our Chinese respondents on a seven-point scale whether they agree or disagree that China or America has won in the conflict. As figure 5 shows, the results are nuanced. In the domestic win-frame scenario where China backs down (Scenario A), respondents were more likely to believe that China has won and the United States has lost when they were exposed to the Chinese win frames (Panels A and B). However, in the foreign win-frame scenario where the United States backs down (Scenario B), the increase in the perception that the United States has won is minimal (Panel D), while there were significant decreases in the perception that China has won (Panel C).

To unpack the asymmetry across the two scenarios, we examined the open responses from respondents who disagreed that China has handled the issue well in the five American win-frame treatment groups (Groups 8–12). Interestingly, only about a quarter made direct references to America winning or China losing.[23] The majority instead inferred from America's win frames that China had failed to actively manage the conflict. A common theme among these respondents is that the US's declaration of "winning" reflects China's failure to take a hardline stance, which, according to one, "indirectly permitted America to have the last laugh." Many even went as far as to assert that China should have continued the trade war to "teach America a lesson," even when China would suffer further costs in doing so. Some respondents also vented their anger about America's declaration of "winning" on the Chinese government:

[21] Results are similar in *t*-tests using the seven-point ordinal scores: relative gains (diff in means = –0.60, $p < 0.001$, $n = 552$), relative losses (diff in means = –0.53, $p < 0.001$, $n = 543$), absolute gains (diff in means = –0.60, $p < 0.001$, $n = 553$), and strategic gains (diff in means = –0.65, $p < 0.001$, $n = 548$), strategic losses (diff in means = –0.86, $p < 0.001$, $n = 543$).

[22] Nonetheless, as discussed below, there are ways for China to control the risk of backlash from foreign win frames.

[23] 193 respondents received win frames about America and disagreed (somewhat disagreed, disagreed, or strongly disagreed) that China had handled the incident well, of which 17 open responses were empty. The remaining 176 responses are coded manually by a native speaker into six categories: (i) the United States gained, (ii) China lost, (iii) the trade war was mutually damaging, (iv) China should have stopped America from declaring victory, (v) China should have been more hardline, and (vi) others. Codings are not mutually exclusive and can overlap.

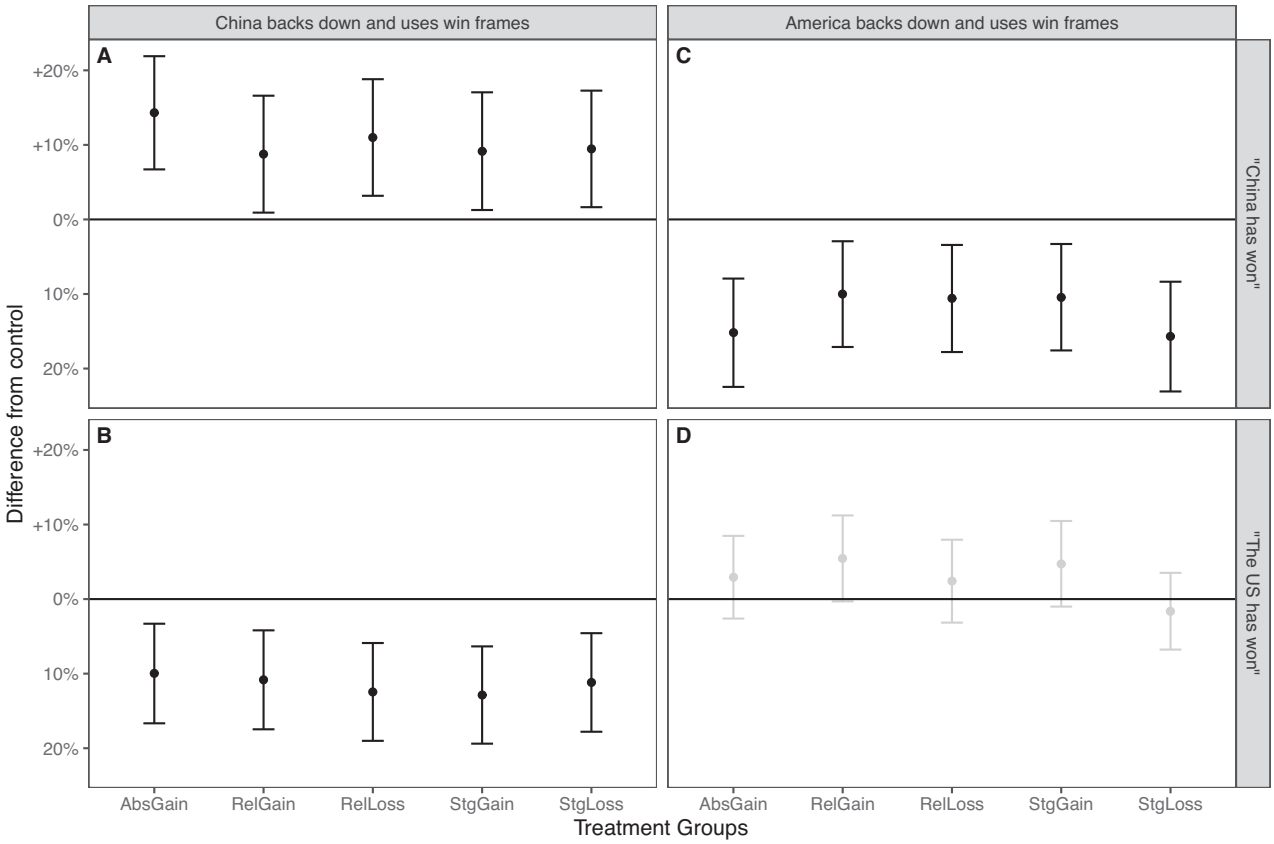

Figure 5. Changes in perception that China or America has won compared to control.
Note: Black solid lines indicate differences with $p < 0.05$. Error bars indicate the 95 percent confidence intervals based on pooled standard errors of the differences in means.

for these respondents, America should not have claimed that it "won" as it was the one that backed down, but China was also culpable for failing to publicly rebut America's "false declaration."

These responses suggest that foreign win frames can hamper the Chinese government from reaping the public approval dividends of the US retreat. However, we see two silver linings here. First, the government may be able to reverse these inferences by counter-framing. It will therefore be interesting for future work to investigate if *counter-frames* to win frames can control the reverberations of foreign win frames. Second, foreign rhetoric in the real world is usually reported in less detail and with domestic audiences paying less attention than domestic rhetoric. In authoritarian countries such as China, the foreign rhetoric may not be reported at all if the government finds it unfavorable. The twin phenomena of asymmetric reporting and asymmetric attention can therefore reinforce the efficacy of win-win deescalation, rendering the mechanism easier for governments to use when they want to back down without "losing."

Finally, to understand the wider implications of the win frames on US–China relations, we asked respondents how they think China's and America's reputations would be impacted following the end of the conflict. This is measured on a seven-point scale from 1 (damaged a lot) to 7 (improved a lot) (for similar measures of reputational impact, see Brutger and Kertzer 2018; Quek 2022). Figure 6 shows the net change in reputation in the treatment groups compared to the control. The public view of China's reputation was *improved* by the Chinese win frames (Panel A) and *worsened* by American win frames (Panel B). This pattern is consistent and statistically significant across all win frames, highlighting that simple rhetorical spins can shape how the Chinese domestic public perceives China's international reputation. On the other hand, Chinese respondents did not think that American win frames improved America's reputation (Panel D), even though those win frames worsened Chinese perception of their own reputation.

We also asked respondents on a scale from 0 (absolutely impossible) to 10 (absolutely possible) about the prospects of the United States and China being friends in the future. The results were assuring: there was no difference in Chinese perception that the United States and China can be friends among all ten treatment groups compared to the control,[24] which suggests that respondents did not see a declaration of "winning" by either party as particularly stigmatic to their future relations.

Our results show that win frames are effective in mitigating the costs of backing down, based on the setting of the US–China trade war. While this is a high-stake and publicly escalated conflict, the trade war is not a militarized conflict. Do win frames also apply to high-intensity milita-

[24]The average score is 5.35 points in the China backdown scenario and 5.44 points in the America backdown scenario.

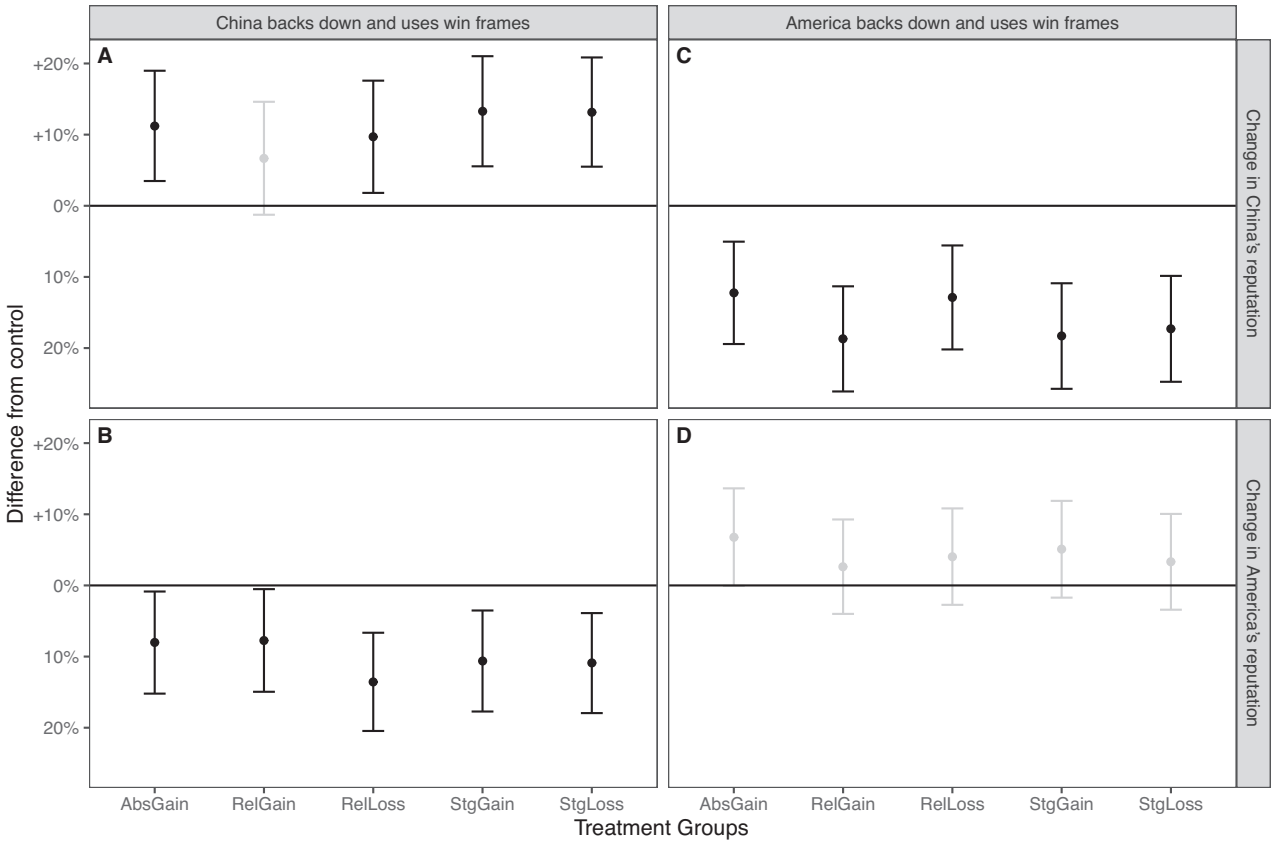

Figure 6. Changes in China's and America's reputation compared to control.
Note: Black solid lines indicate differences with $p < 0.05$. Error bars indicate the 95 percent confidence intervals.

rized conflicts? We propose that they do. Theoretically, so long as there exists uncertainty about the facts on the battleground and/or what it means to win in the conflict, leaders can leverage the uncertainty to shape the narrative (see the section "Backing Down, Framing, and Winning"). Empirically, historical evidence suggests leaders do use win frames in high-stakes militarized conflicts to protect their reputation. Consider the 2019 Balakot crisis, where Indian aircraft crossed into Pakistan to strike an alleged Jaish-e-Mohammed (JeM) training camp, triggering a crisis between the two nuclear-armed powers. The Indian government lauded the success of the operation, claiming that it had destroyed the "biggest training camp of JeM in Balakot" and eliminated "a very large number of JeM terrorists, trainers, senior commanders and groups of jihadis."[25] Pakistan then launched a retaliatory airstrike, which the Pakistan government framed as a success in "demonstration of [Pakistan's] capability, responsibility and will" as it called for the crisis to come to a close.[26] The Balakot crisis is only one of numerous examples in which belligerents deescalate a crisis through the use of win frames. Online appendix E includes three additional case illustrations of how win frames are used in militarized conflicts across different times and places.

Conclusion

Leaders who find themselves in an international dispute are often between a rock and a hard place. Continuing the conflict would incur further costs, yet backing down would be politically costly. This article proposes a mechanism of conflict deescalation that has not been previously theorized and tested: *win-win deescalation*.[27] State A "wins" because State B retreats, yet State B also "wins" because it successfully frames its action to its audience as a "win." Turning to empirics, we show that leaders can control their costs of backing down by deploying win frames—schemata of interpretation that are used to shape the perception of winning and, by extension, political approval. Fielding an experiment in the context of a US–China trade war, we find that the government can finesse the decision to back down by using different win frames to induce the perception of winning and reduce its domestic costs of backing down. Conversely, American win frames can also make it politically harder for the Chinese government to back down if they are publicized to the Chinese public.

Our article has theorized and tested the use of win frames as a mechanism for conflict deescalation. We show that the multidimensional and contestable nature of the idea of winning opens a discursive space for leaders to use win frames

[25]Ministry of External Affairs. 2019. "Statement by Foreign Secretary on 26 February 2019 on the Strike on JeM training camp at Balakot."

[26]"Pakistan wants peace; India needs to understand war is a failure of policy: DG ISPR." Geo TV. February 27, 2019.

[27]There is an additional form of win-win deescalation whereby both States A and B compromise in the conflict and their relevant audience perceives their country has "won" or at least not "lost." Win frames are still relevant nonetheless, as a mutual compromise does not necessarily mean that audiences would perceive it as a mutual "win."

that can render an unsatisfactory international outcome more palatable domestically. Understanding how win frames function would be important given the hardening of domestic sentiments against each other in the United States and in China, increasing the risks of a dangerous confrontation between the two great powers. Furthermore, insofar as the efficacy of win frames hinges on the domestic media environment, authoritarian governments such as China that can control the media would have a clear advantage in deploying win frames domestically to deescalate internationally.

Our article also provides the first experimental evidence on the effectiveness and implications of both domestic and foreign win frames, both from one's own government and from a foreign government. While survey experiments in IR generally focus on the effects of words or actions in one country, our experimental design simultaneously tests the effect of the same win frames deployed by both sides of the dyadic conflict. Through this design, we show that win frames that increase public approval on one side may also decrease domestic approval on the other side. This, of course, assumes that the win frames deployed by one government are publicized in the rival country. Authoritarian governments may be better able to block foreign win frames from being publicized in their own country. Thus, authoritarian governments such as China may have more precise control than democratic governments over the use of win-win deescalation.

While our study focuses empirically on a trade war scenario, its implications can be general. The theoretical logic in our mechanism of win-win deescalation has a general form. In any conflict where leaders enjoy an informational advantage over the public and there is no clear metric of success due to the multidimensionality and contestability of the issue, there is a discursive space that leaders can use to frame their own win. That win frames are effective in reducing domestic backlash among a nationalistic Chinese public—a hard case—also suggests the mechanism's potential efficacy in other countries.

Our work points to several avenues for future exploration. First, our experiment assumes that win frames are directly communicated by the government to the public. This is more realistic in the case of China, where the media is tightly controlled, but less so in liberal democracies where media power is decentralized and fragmented. Future studies may find it fruitful to investigate the effectiveness of win frames in democratic contexts with competitive media markets, especially when these win frames are filtered through partisan media outlets. Second, we have devised and tested a repertoire of win frames based on a zero-sum context that invokes relative gains (losses) and strategic gains (losses). It will be useful for future work to devise and test a greater variety of win frames, including those in the positive-sum context, whereby one's retreat is framed as a "win" for everyone, including the rival state—and possibly also the region or even the world. The slippery meaning of winning also points to the possibility of win frames devised in dimensions that are ideational (moral, normative, etc.) rather than material.[28] A state can frame its retreat as a moral victory or as the correct action of a responsible great power.[29] Finally, our experiment has leveraged the context of an ongoing US–China trade war. While there is observational evidence that win frames are also used by leaders in high-intensity militarized conflicts, a useful avenue for future work is to establish causality more precisely through experiments in the context of salient militarized crises. For instance, it will be useful to do a stress test of the efficacy of win frames in an emotionally charged conflict, such as the conflict between the United States and China over Taiwan.[30] Experimental replication on less emotionally charged issues focusing on other conflict dyads is useful too. We have focused empirically on China and the United States because they are a conflict dyad that is very important to the world, but win-win deescalation is a general mechanism that may apply theoretically to other conflicts as well.

Supplementary Information

Supplementary information is available in the *International Studies Quarterly* data archive.

References

Ajzen, Icek, and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Aksoy, Deniz, Ted Enamorado, and Tony Zirui Yang. 2024. "Russian Invasion of Ukraine and Chinese Public Support for War." *International Organization* 78 (2): 341–60.

Amiti, Mary, Sang Hoon Kong, and David Weinstein. 2020. "The Effect of the U.S.-China Trade War on U.S. Investment." National Bureau of Economic Research, Working Paper 27114.

Angstrom, Jan. 2007. "The United States Perspective on Victory in the War on Terrorism." In *Understanding Victory and Defeat in Contemporary War*, edited by Jan Angstrom and Isabelle Duyvesteyn, 94–113. London: Routledge.

Bali, Valentina. 2007. "Terror and Elections: Lessons from Spain." *Electoral Studies* 26 (3): 669–87.

Bartholomees, J. Boone. 2010. "A Theory of Victory." In *Theory of War and Strategy*, edited by J. Boone Bartholomees, 79–94. Carlisle, PA: Strategic Studies Institute.

Baum, Matthew, and Tim Groeling. 2010. "Reality Asserts Itself: Public Opinion on Iraq and the Elasticity of Reality." *International Organization* 64 (3): 443–79.

Baum, Matthew, and Philip Potter. 2008. "The Relationships between Mass Media, Public Opinion, and Foreign Policy: Toward a Theoretical Synthesis." *Annual Review of Political Science* 11 (1): 39–65.

Bell, Mark, and Kai Quek. 2018. "Authoritarian Public Opinion and the Democratic Peace." *International Organization* 72 (1): 227–42.

Berinsky, Adam. 2007. "Assuming the Costs of War: Events, Elites, and American Public Support for Military Conflict." *The Journal of Politics* 69 (4): 975–97.

Bond, Brian. 1996. *The Pursuit of Victory: From Napoleon to Saddam Hussein*. Oxford: Oxford University Press.

Breslin, Shaun. 2010. "China's Emerging Global Role: Dissatisfied Responsible Great Power." *Politics* 30 (1 suppl): 52–62.

Brutger, Ryan, and Joshua Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72 (3): 693–724.

Burk, James. 1999. "Public Support for Peacekeeping in Lebanon and Somalia: Assessing the Casualties Hypothesis." *Political Science Quarterly* 114 (1): 53–78.

Callahan, William. 2004. "National Insecurities: Humiliation, Salvation, and Chinese Nationalism." *Alternatives: Global, Local, Political* 29 (2): 199–218.

Campbell, James. 2010. "The Exceptional Election of 2008: Performance, Values, and Crisis." *Presidential Studies Quarterly* 40 (2): 225–46.

Carlson, Lisa. 1995. "A Theory of Escalation and International Conflict." *Journal of Conflict Resolution* 39 (3): 511–34.

[28] On how normative framing relates to untying "tied hands," see Quek (2022). See also Goddard's (2018) argument that a rising power can use a legitimizing frame to make its assertive behavior appear less threatening.

[29] On the rise and decline of the idea of China as a responsible great power, see Deng (2014). See also Xia (2001) and Breslin (2010).

[30] Individuals with different beliefs about the indivisibility of the territory (Fang, Li, and Liu 2022) may respond differently to the same win frame in a territorial dispute.

Chan, Zenobia, Samuel Liu, and Kai Quek. 2021. "Audience Costs and the Credibility of Commitments." In *Oxford Bibliographies: International Relations*. Oxford: Oxford University Press.

Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization* 68 (1): 235–56.

Chong, Dennis, and James Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10 (1): 103–26.

Chyi, Hsiang Iris, and Maxwell McCombs. 2004. "Media Salience and the Process of Framing: Coverage of the Columbine School Shootings." *Journalism & Mass Communication Quarterly* 81 (1): 22–35.

Clare, Joe, and Vesna Danilovic. 2012. "Reputation for Resolve, Interests, and Conflict." *Conflict Management and Peace Science* 29 (1): 3–27.

Colaresi, Michael. 2007. "The Benefit of the Doubt: Testing an Informational Theory of the Rally Effect." *International Organization* 61 (1): 99–143.

Copeland, Dale. 1997. "Do Reputations Matter?" *Security Studies* 7 (1): 33–71.

Dafoe, Allan, Samuel Liu, Brian O'Keefe, and Jessica Chen Weiss. 2022. "Provocation, Public Opinion, and International Disputes: Evidence from China." *International Studies Quarterly* 66 (2): sqac006.

Davies, Graeme, and Robert Johns. 2013. "Audience Costs among the British Public: The Impact of Escalation, Crisis Type, and Prime Ministerial Rhetoric." *International Studies Quarterly* 57 (4): 725–37.

de Vreese, Claes. 2004. "The Effects of Frames in Political Television News on Issue Interpretation and Frame Salience." *Journalism & Mass Communication Quarterly* 81 (1): 36–52.

Deng, Yong. 2014. "China: The Post-Responsible Power." *The Washington Quarterly* 37 (4): 117–32.

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65 (2): 135–50.

Druckman, James, and Kjersten Nelson. 2003. "Framing and Deliberation: How Citizens' Conversations Limit Elite Influence." *American Journal of Political Science* 47 (4): 729–45.

Druckman, James, Erik Peterson, and Rune Slothuus. 2013. "How Elite Partisan Polarization Affects Public Opinion Formation." *American Political Science Review* 107 (1): 57–79.

Duan, Xiaolin. 2017. "Unanswered Questions: Why We May Be Wrong about Chinese Nationalism and Its Foreign Policy Implications." *Journal of Contemporary China* 26 (108): 886–900.

Fang, Songying, Xiaojun Li, and Adam Y. Liu. 2022. "Chinese Public Opinion about US–China Relations from Trump to Biden." *The Chinese Journal of International Politics* 15 (1): 27–46.

Fang, Songying, Xiaojun Li, Atsushi Tago, and Daina Chiba. 2022. "Belief in Territorial Indivisibility and Public Preferences for Dispute Resolution." *Political Science Research and Methods* 10 (4): 759–75.

Fang, Songying, and Xiaojun Li. 2020. "Historical Ownership and Territorial Disputes." *The Journal of Politics* 82 (1): 345–60.

Fearon, James. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92.

Feaver, Peter, and Christopher Gelpi. 2011. *Choosing Your Battles: American Civil-Military Relations and the Use of Force*. Princeton, NJ: Princeton University Press.

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

Gartner, Scott Sigmund, and Gary Segura. 1998. "War, Casualties, and Public Opinion." *Journal of Conflict Resolution* 42 (3): 278–300.

Gelpi, Christopher, and Joseph Grieco. 2015. "Competency Costs in Foreign Affairs: Presidential Performance in International Conflicts and Domestic Legislative Success, 1953–2001." *American Journal of Political Science* 59 (2): 440–56.

Gilboa, Eytan. 2005. "The CNN Effect: The Search for a Communication Theory of International Relations." *Political Communication* 22 (1): 27–44.

Goddard, Stacie. 2018. *When Right Makes Might*. Ithaca, NY: Cornell University Press.

Goffman, Erving. 1986 [1974]. *Frame Analysis: An Essay on the Organization of Experience*. Boston, MA: Northeastern University Press.

Grieco, Joseph. 1988. "Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism." *International Organization* 42 (3): 485–507.

Gries, Peter, and Tao Wang. 2020. "Taiwan's Perilous Futures: Chinese Nationalism, the 2020 Presidential Elections, and U.S.-China Tensions Spell Trouble for Cross-Strait Relations." *World Affairs* 183 (1): 40–61.

Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (2): 175–200.

Hiscox, Michael. 2006. "Through a Glass and Darkly: Attitudes toward International Trade and the Curious Effects of Issue Framing." *International Organization* 60 (3): 755–80.

Hobolt, Sara, and Catherine de Vries. 2016. "Turning against the Union? The Impact of the Crisis on the Eurosceptic Vote in the 2014 European Parliament Elections." *Electoral Studies* 44: 504–14.

Jervis, Robert. 1997. *System Effects: Complexity in Political and Social Life*. Princeton, NJ: Princeton University Press.

———. 2002. "Signaling and Perception." In *How Statesmen Think: The Psychology of International Politics*, edited by Robert Jervis. Princeton, NJ: Princeton University Press.

Johnson, Dominic, and Dominic Tierney. 2006. *Failing to Win: Perceptions of Victory and Defeat in International Politics*. Cambridge, MA: Harvard University Press.

———. 2007. "In the Eye of the Beholder." In *Understanding Victory and Defeat in Contemporary War*, edited by Jan Angstrom and Isabelle Duyvesteyn, 46–76. London: Routledge.

Kertzer, Joshua, Ryan Brutger, and Kai Quek. 2024. "Perspective Taking and the Security Dilemma: Cross-National Experimental Evidence from China and the United States." *World Politics* 76 (2): 334–78.

Kertzer, Joshua, and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60 (1): 234–49.

Kertzer, Joshua. 2019. *Resolve in International Politics*. Princeton, NJ: Princeton University Press.

Kim, Il Sung. 1981. *Kim Il Sung Works, Volume 7, January 1952–July 1953*. Pyongyang: Foreign Languages Publishing House.

King, Gary, Jennifer Pan, and Margaret Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107 (2): 326–43.

Ko, Jiyoung. 2022. "Not So Dangerous? Nationalism and Foreign Policy Preference." *International Studies Quarterly* 66 (3): sqac053.

Kohama, Shoko, Kai Quek, and Atsushi Tago. 2024. "Managing the Costs of Backing Down: A 'Mirror Experiment' on Reputations and Audience Costs in a Real-World Conflict." *The Journal of Politics* 86 (1): 388–93.

Kuang, Xianwen. 2023. "Media Type and the Framing of the Sino-US Trade War: An Analysis of Articles from Party and Nonparty News Organisations in China." In *China, Media, and International Conflicts*, edited by Shixin Ivy Zhang and Altman Yuzhu Peng. London: Routledge.

Kuran, Timur. 1998. *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.

Levendusky, Matthew, and Michael Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74 (2): 323–38.

Levy, Jack, Michael McKoy, Paul Poast, and Geoffrey Wallace. 2015. "Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59 (4): 988–1001.

Li, Xiaojun, and Dingding Chen. 2021. "Public Opinion, International Reputation, and Audience Costs in an Authoritarian Regime." *Conflict Management and Peace Science* 38 (5): 543–60.

Lin-Greenberg, Erik. 2019. "Backing Up, Not Backing Down: Mitigating Audience Costs through Policy Substitution." *Journal of Peace Research* 56 (4): 559–74.

Lupia, Arthur. 1994. "Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review* 88 (1): 63–76.

Lupton, Danielle. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca, NY: Cornell University Press.

Mandel, Robert. 2006. *The Meaning of Military Victory*. Boulder, CO: Lynne Rienner Publishers.

———. 2007. "Defining Postwar Victory." In *Understanding Victory and Defeat in Contemporary War*, edited by Jan Angstrom and Isabelle Duyvesteyn, 13–45. London: Routledge.

Martel, William. 2007. *Victory in War: Foundations of Modern Military Policy*. Cambridge: Cambridge University Press.

McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.

McManus, Roseanne. 2018. "Making It Personal: The Role of Leader-Specific Signals in Extended Deterrence." *The Journal of Politics* 80 (3): 982–95.
Mueller, John. 1973. *War, Presidents, and Public Opinion*. New York, NY: Wiley.
Myrick, Rachel, and Chen Wang. 2024. "Domestic Polarization and International Rivalry: How Adversaries Respond to America's Partisan Politics." *The Journal of Politics* 86 (1): 141–57.
Nalebuff, Barry. 1991. "Rational Deterrence in an Imperfect World." *World Politics* 43 (3): 313–35.
Nelson, Thomas, Zoe Oxley, and Rosalee Clawson. 1997. "Toward a Psychology of Framing Effects." *Political Behavior* 19 (3): 221–46.
Ong, Lynette, and Donglin Han. 2019. "What Drives People to Protest in an Authoritarian Country? Resources and Rewards vs Risks of Protests in Urban and Rural China." *Political Studies* 67 (1): 224–48.
Paterson, Thomas, and William Brophy. 1986. "October Missiles and November Elections: The Cuban Missile Crisis and American Politics, 1962." *The Journal of American History* 73 (1): 87–119.
Peterson, Erik, and Gabor Simonovits. 2018. "The Electoral Consequences of Issue Frames." *The Journal of Politics* 80 (4): 1283–96.
Potter, Philip, and Chen Wang. 2023. *Zero Tolerance: Repression and Political Violence on China's New Silk Road*. New York, NY: Cambridge University Press.
Powell, Robert. 1991. "Absolute and Relative Gains in International Relations Theory." *American Political Science Review* 85 (4): 1303–20.
Quek, Kai, and Alastair Iain Johnston. 2018. "Can China Back Down? Crisis De-Escalation in the Shadow of Popular Opposition." *International Security* 42 (3): 7–36.
Quek, Kai. 2017. "Type II Audience Costs." *The Journal of Politics* 79 (4): 1438–43.
———. 2022. "Untying Hands: De-Escalation, Reputation, and Dynamic Audience Costs." *British Journal of Political Science* 52 (4): 1964–76.
Riker, William. 1990. "Heresthetic and Rhetoric in the Spatial Model." In *Advances in the Spatial Theory of Voting*, edited by James M. Enelow and Melvin J. Hinich, 46–65. Cambridge: Cambridge University Press.
Roberts, Margaret. 2018. *Censored: Distraction and Diversion inside China's Great Firewall*. Princeton, NJ: Princeton University Press.
Robinson, Piers. 1999. "The CNN Effect: Can the News Media Drive Foreign Policy?" *Review of International Studies* 25 (2): 301–9.
Rousseau, David. 2002. "Motivations for Choice: The Salience of Relative Gains in International Politics." *Journal of Conflict Resolution* 46 (3): 394–426.
Sartori, Anne. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (1): 121–49.
Schultz, Kenneth. 2001. *Democracy and Coercive Diplomacy*. New York, NY: Cambridge University Press.
Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
Snidal, Duncan. 1991. "Relative Gains and the Pattern of International Cooperation." *American Political Science Review* 85 (3): 701–26.
Sniderman, Paul, Richard Brody, and Philip Tetlock, eds. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge: Cambridge University Press.
Sniderman, Paul, and Sean Theriault. 2004. "The Structure of Political Argument and the Logic of Issue Framing." In *Studies in Public Opinion*, edited by Willem Saris and Paul Sniderman, 133–65. Princeton, NJ: Princeton University Press.
Sniderman, Paul. 1993. "The New Look in Public Opinion Research." In *Political Science: The State of the Discipline II*, edited by Ada W. Finifter, 219–45. Washington, DC: American Political Science Association.
Snyder, Glenn Herald, and Paul Diesing. 1977. *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, NJ: Princeton University Press.
Stockmann, Daniela. 2013. *Media Commercialization and Authoritarian Rule in China*. Cambridge: Cambridge University Press.
Takei, Makito, and Philip Paolino. 2023. "Backing Out but Backing in Audience Costs? A Replication of Levy et al. (2015)." *Foreign Policy Analysis* 19 (2): orad008.
Tobin, Jennifer, Christina Schneider, and David Leblang. 2022. "Framing Unpopular Foreign Policies." *American Journal of Political Science* 66 (4): 947–60.
Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–40.
Trager, Robert, and Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal of Political Science* 55 (3): 526–45.
Wallace, Jeremy, and Jessica Chen Weiss. 2015. "The Political Geography of Nationalist Protest in China: Cities and the 2012 Anti-Japanese Protests." *The China Quarterly* 222: 403–29.
Waltz, Kenneth Neal. 1959. *Man, the State and War: A Theoretical Analysis*. New York, NY: Columbia University Press.
———. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.
Wang, Zheng. 2008. "National Humiliation, History Education, and the Politics of Historical Memory: Patriotic Education Campaign in China." *International Studies Quarterly* 52 (4): 783–806.
Weeks, Jessica. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35–64.
Weiss, Jessica Chen, and Allan Dafoe. 2019. "Authoritarian Audiences, Rhetoric, and Propaganda in International Crises: Evidence from China." *International Studies Quarterly* 63 (4): 963–73.
Weiss, Jessica Chen. 2014. *Powerful Patriots: Nationalist Protest in China's Foreign Relations*. New York, NY: Oxford University Press.
Wolford, Scott. 2014. "Showing Restraint, Signaling Resolve: Coalitions, Cooperation, and Crisis Bargaining." *American Journal of Political Science* 58 (1): 144–56.
Xia, Liping. 2001. "China: A Responsible Great Power." *Journal of Contemporary China* 10 (26): 17–25.
Yeung, Eddy, and Kai Quek. 2022. "Relative Gains in the Shadow of a Trade War." *International Organization* 76 (3): 741–65.
Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Koo, John N. L., and Kai Quek. (2024) Win-Win Deescalation. *International Studies Quarterly*, https://doi.org/10.1093/isq/sqae148