Perspective Taking and Security Dilemma Thinking:
Experimental Evidence from China and the United States

Joshua D. Kertzer,¹ Ryan Brutger,² and Kai Quek³

October 1, 2023

World Politics, Forthcoming.

Abstract: One of the central challenges in China-US relations is the risk of a security dilemma between China and the United States, as each side carries out actions for what it perceives to be defensively-motivated reasons, failing to realize how it is perceived by the other side. Yet how susceptible to the psychological biases that undergird the security dilemma are the Chinese and American publics? Can their deleterious effects be mitigated? We explore the microfoundations of the security dilemma, fielding parallel dyadic cross-national survey experiments in China and the United States. We find micro-level evidence consistent with the logic of the security dilemma in publics in both countries. We also find that IR scholars have overstated the palliative effects of perspective taking, which can backfire in the face of perceived threats to actors’ identities and goals. Our findings have important implications for the study of public opinion in China-US relations, and perspective taking in IR.

Acknowledgments: We’re grateful to Michael Chen, Thomas Chung, Briggs DeLoach, Hope Ha, Jiaqian Ni, Samuel Liu, Nic Pantelick, Felix Wang, Xiao Han, and Eddy Yeung for research assistance, Chris Carothers, Rush Doshi, Austin Strange, and George Yin for their help with the Chinese survey instrument, and Mark Bell, Matt Blackwell, Emanuele Castano, Cosette Creamer, Tanisha Fazal, Peter Gries, Joshua Gubler, Rick Herrmann, Iain Johnston, Ron Krebs, Marika Landau-Wells, Howie Lavine, Matt Motta, Rich Petty, Lauren Prather, Brian Rathbun, Margit Tavits, and audiences at APSA, ISA, Laval, Minnesota, UVA, and Ohio State for helpful feedback. Kertzer acknowledges the support of the Weatherhead Center for International Affairs; Brutger acknowledges the support of the Niehaus Center for Globalization and Governance and the Center for International Security Studies at Princeton University; Quek acknowledges support from the University of Hong Kong.

¹John Zwaanstra Professor of International Studies and of Government, Department of Government, Harvard University. Email: jkertzer@gov.harvard.edu. Web: https://jkertzer.sites.fas.harvard.edu/
²Associate Professor, Travers Department of Political Science, University of California Berkeley. Email: brutger@berkeley.edu. Web: https://sites.google.com/berkeley.edu/brutger/
³Associate Professor, Department of Politics and Public Administration, The University of Hong Kong. Email: quek@hku.hk. Web: https://www.ppaweb.hku.hk/f/quek
Peacefully managing the rise of China is the most pressing foreign policy challenge of the 21st century (Mearsheimer, 2001; Christensen, 2006). Scholars and policymakers who are worried about the risks associated with China’s rise are worried about two phenomena. The first involves the hegemonic wars envisioned by power transition theory (Gilpin, 1981; Kugler and Lemke, 2000; Allison, 2017). The second involves conflicts that can occur between two security-seeking states: the notion of a security dilemma (Herz, 1950; Butterfield, 1951).

Central to the idea of a security dilemma is a dynamic in which one state carries out actions for defensively-motivated reasons, but fails to appreciate how it will be perceived by the other side, leading to a spiral model of conflict that no actor actually wants (Jervis, 1978; Glaser, 1997; Booth and Wheeler, 2008; Tang, 2009; Mitzen and Schweller, 2011). Security dilemmas are a byproduct of the uncertainty generated by the anarchic structure of the international system, but they are also about failures of perspective taking — the inability of one actor to successfully put itself in the shoes of the other.

In this paper, we experimentally explore the psychological microfoundations of the security dilemma. We are interested in two questions. First of all, in an era when pundits routinely express concern about publics in China and the US being on a collision course with one another, how susceptible are citizens in the two countries to the attribution asymmetries that characterize the security dilemma, as individuals attribute offensive motivations to others for behavior they would perceive as defensively motivated if carried out by their own side? Second, given IR scholars’ persistent interest in the palliative effects of empathy and perspective taking in promoting international cooperation (e.g. Jervis, 1976; White, 1986; Keller and Yang, 2009; Holmes and Yarhi-Milo, 2017), what role can perspective taking play in dampening security dilemma dynamics? If security dilemmas are about failures of empathy, does inducing perspective taking and encouraging citizens from different countries to step into each other’s shoes mitigate security dilemma thinking’s deleterious effects?

We explore these questions using a series of cross-national survey experiments fielded in both China and the United States, where we present participants in both countries with a scenario in the South China Sea, manipulate the behavior of the other side, and examine its effects on policy preferences. We find micro-level evidence consistent with the attributional asymmetry at the heart of psychological theories of the security dilemma: respondents in both countries are susceptible to spiral models of conflict, viewing behavior as offensively-motivated when carried out by the other side, and defensively-motivated when carried out by their own – particularly among respondents
in China. Yet we also find that perspective taking is a double-edged sword, because its effects are conditional on the knowledge structures it activates: experimentally inducing American participants to think about the conclusions the Chinese would draw from their behavior increases the likelihood of endorsing a policy of reciprocity, but doing the same for Chinese participants can actually increase the likelihood of endorsing escalation, due to differences in threat perception by each side. Our findings have important implications for both theory and policy, improving our understanding of the logic of the security dilemma and the role of the mass public in either accelerating or inhibiting assertive foreign policies in Chinese-US relations, and adding a cautionary note to the conventional wisdom about the palliative effects of perspective taking in foreign affairs.

The discussion that follows has four parts. We begin by reviewing the security dilemma, a classic framework in IR whose resonance has only increased in light of growing tensions in the South China Sea. We then show how psychological models of the security dilemma carve out a particularly important role for perspective-taking, a cognitive form of empathy that plays a crucial role in strategic behavior more generally. Although IR scholars tend to emphasize perspective taking’s palliative effects, it can also have a dark side, particularly when actors perceive threats to their identities or goals. We then discuss the mechanics of our cross-national experiments, before presenting our findings. Finally we conclude with implications for theory and policy.

1 The security dilemma in US-China relations

Policymakers and political scientists preoccupied with the predicament posed by a rising China (e.g. Christensen, 1999; Mearsheimer, 2001; Johnston, 2003; Medeiros and Fravel, 2003; Friedberg, 2005; Gries, 2005; Goldstein, 2005; Christensen, 2006; Ikenberry, 2008; Ross, 2009; Schweller and Pu, 2011; Chen, Pu and Johnston, 2014; Johnston, 2017) typically focus on one of two phenomena.1

The first is the specter of hegemonic war. As Gilpin (1981) argued, the international political order tends to reflect the interest of the most powerful states in the system; as the balance of power shifts due to the law of uneven growth, the newly powerful will attempt to change the system to better reflect their interests. Declining hegemons, then, have an incentive to wage preventive war in order to forestall the rising power’s rise (Levy, 1987), such that periods of power transition have

---

1The 2012 TRIP survey suggests the strategic importance of peacefully managing China’s rise is shared by both academics and policymakers alike, topping the list of foreign-policy problems facing the United States provided by both IR scholars in the United States, and national security practitioners within the US government. See https://www.wm.edu/offices/stpir/_documents/trip/ivory_tower_survey_2012.pdf.
historically tended to be periods of instability (Kugler and Lemke, 2000; Allison, 2017). Importantly, however, the dire predictions of power transition theory are contingent upon the rising power being a revisionist state, willing to pay a high cost to overturn the established order.\(^2\) If China is invested in the existing international political order — an arrangement that enabled it to lift nearly three quarters of a billion people out of poverty over the past several decades — then no conflict should occur.

The second is the security dilemma (Herz, 1950; Butterfield, 1951). As Booth and Wheeler (2008, 4-8) and Tang (2009) note, IR scholars use the term in a wide variety of contradictory ways, but often beginning with the premise that because of the anarchic structure of the international system, states are fundamentally uncertain about the intentions of others (Waltz, 1979; Copeland, 2000; Mearsheimer, 2001). When one state takes actions that are intended only to bolster its own security, other states are likely to assume the worst, and erroneously interpret a defensively-motivated action as an offensively-motivated one. The result is what Jervis (1976) calls the “spiral model” of conflict, where war no one wants emerges as a result of misperceptions that neither side can shake off (Glaser, 1997; Tang, 2009; Mitzen and Schweller, 2011). Thus, whereas hegemonic war only occurs when one of the actors is a revisionist state, the security dilemma can occur even among two states supportive of the status quo. Concerns about potential security dilemmas in East Asia — whether between China and the United States, or between China and America’s allies in the region — thus loom large in much of the East Asian security scholarship (e.g. Christensen, 1999, 2002; Breuer and Johnston, 2019), with Western scholars warning of a “growing security dilemma which could spiral into a regional arms race, destabilizing Asia and increasing the chance of conflict if there is not a swift shift in direction” (Ludwig, 2017). Yet Chinese scholars have recently made similar arguments, suggesting that “the South China Sea dispute has become the ‘security dilemma’ of the two countries” (南海成为两国的“安全困境”), or looking for ways for the US and China to “step over the trap of ‘security dilemmas’ in the South China Sea” (并进而跨越南海“安全困境”的陷阱).\(^3\)

What unites these two worst-case scenarios is the challenge of assessing intentions (Yarhi-Milo, 2014). Because hegemonic wars depend on the rising power harboring revisionist intentions, debates

---

\(^2\)As Jervis (1976, 50), what makes actors revisionist is not whether they have unsatisfied goals (all states do), but how high the costs they are willing to pay in order to achieve them.

in Washington over how the United States should handle a rising China — whether we should engage China and enmesh it in international institutions (Ikenberry, 2008), reassure China by using reciprocal gestures to reduce tensions (Steinberg and O’Hanlon, 2014), contain it militarily (Mearsheimer, 2014), or some combination thereof — are partially debates about future states of the world (e.g. how much will Chinese economic growth slow? How much will Chinese nationalism rise?), but are also debates about what China wants (Johnston, 2003). Ultimately, American grand strategy debates over the Asia-Pacific hinge on the fundamental challenge of overcoming “the problem of other minds” (Herrmann, 1988) and accurately assessing the motivations of others. As Butterfield (1951, 154) noted when coining what he referred to as the “irreducible dilemma”, these same issues serve as the core of the security dilemma as well.

In his influential discussion of the logic of spiral models and security dilemmas, Jervis (1976, 68-69) points to two different psychological dynamics we see as central psychological microfoundations for the security dilemma. First, actors attribute offensive motivations to others for behavior they would perceive as defensively motivated if carried out by their own side (Jervis, 1976, 68). We call this psychological dynamic an attribution asymmetry, since it refers to a double standard in the motivations we assign to others’ behavior compared to our own (Canetti, Gubler and Zeitzoff, 2021). As a result of the attribution asymmetry, actors on each side espouse “benign memes about the Self and malign memes about the Other” (Breuer and Johnston, 2019, 432), and become more likely to want to escalate, leading to a potential spiral of conflict as each side responds to one another’s actions, blaming their escalation on the other side. Johnston (2003, 50) notes that this interactivity can even take the form of “malign reciprocation”, where “conflictual moves reciprocate both conflictual and cooperative actions.” In the first experiment we present below, we seek to provide micro-level evidence consistent with psychological models of the security dilemma by looking for attribution asymmetries among the American and Chinese publics. In the second experiment, we test the efficacy of one frequently touted solution to attribution asymmetries by turning to the second psychological dynamic Jervis (1976, 69) points to in his discussions of security dilemmas: perspective taking.

---

4See Kydd (1997) for a rational choice treatment of the attribution asymmetry.
1.1 The palliative effect of perspective taking

Perspective taking is a psychological faculty that involves the ability or willingness to put oneself in others’ shoes. Because of its centrality to theories of symbolic interactionism, perspective taking is often associated with constructivist scholarship (e.g. Wendt, 1999, 333), but perspective taking is fundamental to all theories of strategic behavior: because strategic situations are those where “the best course of action for each player depends on what other players do” (Schelling, 1960, 3, fn. 1), acting strategically requires anticipating the behavior of others, which requires the capacity to see the situation from someone else’s eyes. Indeed, Singer and Fehr (2005, 340) note that all of “the most fundamental solution concepts in game theory (Nash equilibrium, backward induction, and iterated elimination of dominated strategies)” depend on some form on this faculty.

In games of complete information, perspective taking is trivial: every player already knows what the other players want and believe as a matter of design. In natural settings, however, there are a number of obstacles, most fundamental of which is the “problem of other minds”: our inability to directly access the mental states of others (Morgenthau, 1985; Herrmann, 1988). In international politics, the problem is exacerbated by the anarchic structure of the international system, in which states are fundamentally uncertain about the intentions of others, who have incentives to misrepresent their private information, as highlighted by a robust body of scholarship on the bargaining model of war (Powell, 1987; Morrow, 1989; Fearon, 1995; Ramsay, 2017).

It should not be surprising, then, that many unsavory outcomes in international politics are frequently chalked up to failures of actors to put themselves in others’ shoes, particularly in the voluminous literature on misperceptions in international politics. Stein (1988, 249-250) notes that threat perceptions are hamstrung by a “lack of empathy”, in that “political leaders often display no sensitivity to their adversary’s sense of vulnerability while they dwell heavily on their own perception of threat.” Robert McNamara offered a similar explanation for American missteps in the Vietnam War (Blight and Lang, 2004), as does Smith (2004) in regard to the unilateralism of the 2003 Iraq War. Keller and Yang (2009, 181) pin the 1991 Persian Gulf War on insufficient perspective taking, manifested in the Bush administration failing to appreciate Saddam Hussein’s domestic political

---

5 Perspective taking is often understood as a cognitive form of empathy, such that some political scientists write of “strategic empathy” (Shore, 2014; Waldman, 2014), although empathy is used in a wide variety of ways in political science (e.g. Keohane, 1984; Head, 2016), and the relationship between empathy and perspective taking varies in the psychological literature: in some work, perspective taking is seen as a cause of empathy (Decety, 2005), while others see perspective taking as a specific subtype of empathy (Davis, 1980), or vice versa (Mohr et al., 2013). We sidestep these debates for our purposes, apart from noting the connections between the two constructs.
constraints. Holmes and Yarhi-Milo (2017, 115-116) attribute the failure of the second Camp David summit to Clinton’s failure “to build relational empathy” between Yasser Arafat and Ehud Barak.

Just as IR scholars have tended to attribute miscalculations and undesirable outcomes to insufficient perspective taking, they frequently link perspective taking to positive or prosocial behavior. This is the logic underlying intergroup contact theory (Pettigrew, 1998), one of the motivations behind foreign exchanges (Atkinson, 2010), and a variety of conflict resolution mechanisms (Batson and Ahmad, 2009), including prejudice-reduction interventions (Paluck, 2010; Adida, Lo and Platas, 2018). Holmes (2018) and Holmes and Yarhi-Milo (2017) argue face-to-face diplomacy is helpful because it better permits negotiators to reach mutual understanding; one of the frequently-issued arguments in favor of area studies expertise in policymaking is that a wealth of local knowledge helps decision-makers better take the other side’s perspective (White, 1986; Waldman, 2014).

These positive effects are often posited to extend to the security dilemma. In his influential discussions of the security dilemma, Jervis (1978, 181) discusses environmental features that make security dilemmas more likely to occur, but also explicitly attributes security dilemmas to “failures of empathy”, and notes that “empathy and skillful statesmanship can reduce this danger” (p. 212). An ability to put oneself in others’ shoes is also central to what Booth and Wheeler (2008) call “security dilemma sensibility”, crucial to breaking out of the dilemma. Most recently, Baker (2019, 2) argues that an ability to place oneself in the shoes of others is “pivotal to the de-escalation of security dilemma dynamics.” The logic of these arguments about the palliative effects of perspective taking are straightforward: if the security dilemma is a failure of perspective taking, engaging in perspective taking will cause actors to better appreciate the effects their actions have on others, and be less likely to assume others’ actions are offensively motivated.

### 1.2 The dark side of perspective taking

Although the IR literature tends to view perspective taking through panglossian lenses, the psychological literature’s findings are more mixed. One concern, which we discuss more in Appendix §4.3, concerns questions of accuracy: most of us aren’t very good at perspective taking, and a propensity or willingness to put oneself in others’ shoes doesn’t mean the inferences drawn will be correct. Most relevant for our purposes, however, is that perspective taking also has a dark side

---

6Perspective taking is one of two solutions frequently offered to the security dilemma, the other being costly signals of reassurance (Kydd, 2005).
(Sassenrath, Hodges and Pfattheicher, 2016). Cognitively, perspective taking causes us to activate knowledge structures both about ourselves and the target whose perspective we are taking (Ku, Wang and Galinsky, 2010), such that the effect of perspective taking depends on the content of the knowledge structures being activated (Vorauer, 2013). In particular, perspective taking can backfire and lead to conflict rather than cooperation when mixed with two ingredients often present in IR.\(^7\)

The first is *threats to one’s identity*. When actors engage in perspective taking, they engage in self-other overlap, in which they both project more of themselves onto the target, and see more of the target in themselves (Sassenrath, Hodges and Pfattheicher, 2016). When actors with strong emotional attachment to their ingroup take the perspective of an outgroup member, it can therefore constitute a threat to their social identity, leading to outgroup derogation (Tarrant, Calitri and Weston, 2012). Perspective taking’s effects should therefore be more negative when nationalism is thrown into the mix, as individuals who derive self-esteem from their membership in the national community attempt to take the perspective of the outgroup. These effects should be further exacerbated when taking the perspective not just of an individual member of the outgroup — as is often the case in prejudice-reduction interventions — but of the outgroup as a whole (Barth and Stürmer, 2016).

The second is *threats to one’s goals*. A growing literature in social psychology finds that perspective taking can lead to more negative attitudes or behavior when actors take the perspective of a group seen as threatening their motivations or goals, as in circumstances where actors perceive negative goal interdependence with the outgroup (Okimoto and Wenzel, 2011; Mooijman and Stern, 2016; Sassenrath, Hodges and Pfattheicher, 2016). If perspective taking is an inherent part of strategy, it is because considering the potential reactions of others is part of defining the strategic environment one faces. It does not, however, transform the underlying nature of that environment; it can illuminate the existence of complementary interests where they exist, but cannot transform conflicting interests into harmony (Axelrod and Keohane, 1985). These findings are consistent with a rational choice perspective as well: perspective taking may help players arrive at equilibrium outcomes, but should not change what those equilibria are.\(^8\) In an innovative set of experiments, Caruso, Epley and Bazerman (2006) (see also Epley, Caruso and Bazerman, 2006) studied perspective taking in

\(^7\)Consistent with the dark side of empathy, in American politics empathic concern has been found to increase partisan bias and promote outgroup hostility, and perspective taking does not ameliorate such partisan biases (Simas, Clifford and Kirkland, 2019).

\(^8\)Of course, as we show below, arriving at equilibria becomes more challenging when the players do not share a common understanding of the game they are playing.
the context of two classic games in the social sciences: the prisoners’ dilemma, and trust games. In the prisoners’ dilemma experiment, participants in the control condition were administered the classic prisoners’ dilemma game, where participants are shown a payoff matrix and asked whether they would like to cooperate or defect (the dominant strategy, but which leaves players collectively worse off). Participants in the treatment condition completed the same task, but were also given a perspective taking manipulation that encouraged them to first consider the thoughts of their opponent; participants who were encouraged to engage in perspective taking were twice as likely to defect as those in the control. In the trust game, participants in the control condition were administered the standard trust game from behavioral economics, in which participants are given a sum of money, and have to choose how much to allocate to another player, who invests the money and then decides how much of the proceeds to return to the original player; the initial amount allocated is used as a measure of trust. When a perspective taking treatment was added, players were significantly less trusting than in the control. Pierce et al. (2013, 1986) similarly find that in competitive contexts, “perspective taking is akin to pouring gasoline on a fire”, effectively transforming “the Golden Rule from ‘do unto others as you would have them do unto you’ to ‘do unto others as you think they will try to do unto you.” In contexts where actors are seen as having opposing goals — if actors perceive themselves as playing a game of chicken, rather than an assurance game (Jervis, 1978; Kydd, 1997), for example — perspective taking can therefore make cooperation less likely, by heightening our awareness of conflicts of interest. In an IR context, since threat perceptions are a function of both perceived capabilities and intentions (e.g. Herrmann and Fischerkeller, 1995), this tendency for perspective taking to have negative effects in situations of perceived negative goal interdependence is particularly likely to occur when the target is seen as significantly stronger in capabilities, wherein the threats they pose to an actor’s goals will be seen as more severe.

Importantly, our theory of perspective taking’s contingent effects acknowledges the role of both cognition and context. If the objects under a dispute have different meanings to different actors, perspective taking will activate different knowledge structures, leading to different effects. Yet it also does so without falling prey to essentialism: unlike theories from cross-cultural psychology that posit that Westerners and Asians inherently engage in perspective taking differently (Wu and Keysar, 2007), our theory predicts that perspective taking should have similar effects across groups when perceivers on both sides define the situation in similar ways.
2 Methods

To test for psychological microfoundations for the security dilemma, we conducted a pair of cross-national survey experiments in China and the US in the spring of 2016, in which we presented an identical set of experimental modules to diverse national samples targeted to match census demographics in both countries at the same time. In this sense, our interest is in studying psychological dynamics in the mass public, rather than elite decision-makers (Kertzer, 2022). In experimental research, Tingley (2017) and Gries and Jing (2019) find that many members of the American public tend to think of a rising China using the same commitment problem framing that underlies models of preventive war. Understanding how susceptible the US public is to the attributional asymmetry at the heart of psychological models of the security dilemma — and thereby offering the first experimental study of the security dilemma’s microfoundations — is a similarly worthy endeavor.

Yet understanding whether similar patterns exist in Chinese public opinion is also important. One major concern among Western pundits is that the Chinese Communist Party’s domestic legitimacy is tied to maintaining high levels of economic growth; as growth rates inevitably slow, the regime may instead fan the flames of nationalism in the mass public, redirecting discontent away from Beijing and towards foreign actors like the United States (Friedberg, 2005, 29-30). It is for this reason that outside observers have been so fixated on nationalist sentiment in the Chinese public (Weiss, 2014; Johnston, 2017), particularly if the Chinese government cultivates nationalist protests at home as a way to tie its hands in international negotiations abroad (Gries, 2001; Weiss, 2013). Ross (2009), for example, argues that China’s maritime buildup is attributable to “naval nationalism” in a public that associates great power status with maritime grandeur. More generally, a recent strand of research on authoritarian accountability suggests that the Chinese government is sensitive to public sentiment in a variety of ways (e.g. Chen, Pan and Xu, 2016; Distelhorst and Hou, 2017; Truex, 2016; Meng, Pan and Ping, 2017). Indeed, despite maintaining an authoritarian regime, the Chinese government spends considerable effort and resources monitoring public opinion (舆情) (Batke and Ohlberg, 2020); Weiss and Dafoe (2019) note that the Chinese government employs two million analysts to monitor public sentiment on Chinese social media, precisely because it is aware of the outsized role that domestic sentiments can play in response to international disputes. Better understanding the dynamics of Chinese public opinion towards the use of force (e.g. Reilly, 2011; Li, Wang and Chen, 2016; Weiss and Dafoe, 2019; Bell and Quack, 2018), is thus a politically
consequential task.

In the United States, we fielded our study in English, embedding it in a survey on a national sample of 1,820 American adults, stratified based on census targets for gender, age, household income, and education. In China, we fielded our study in Chinese, embedding it in a national survey that was fielded online to cover all provinces and capital municipalities in Mainland China. Anonymized online surveys are known to reduce social desirability biases and improve response validity, which is particularly important in sensitive environments such as China (e.g. Chang and Krosnick, 2010; Huang, 2015, 630). To conduct our study in China, we partnered with a survey company to recruit a sample of 1,556 Chinese adults (18 years and older) seeking to match the 2010 National Census on gender, age, race, income and geography. Because these subjects were directed to the Qualtrics survey platform at the researcher end, we maintained full control over our experiment and the data collection.

Respondents in both countries were presented with parallel experimental scenarios, involving tensions in the South China Sea between the United States and China. The overall structure of our study is presented in Figure 1, and the Chinese and English text used in our experiments are reproduced alongside each other in Appendix §3. Respondents in both countries began the experimental portion of the study by reading: “Recently there has been much attention over tensions in the South China Sea. Multiple countries in the region have claimed rights to disputed international waters, which are home to a wealth of natural resources, fisheries, and trade routes — all of which are at stake in the increasingly frequent diplomatic standoffs. [China/The United States] is concerned about the [United States'/China's] assertiveness in the region.” That is to say, respondents in China were told about the United States, while respondents in the United States were told about China.

Respondents then completed two experimental modules, which we refer to as the attribution experiment, and the perspective taking experiment, each of which focuses on the two psychological microfoundations for the security dilemma identified in Jervis (1976) and discussed above. The attribution experiment has a simple within-subjects design, in which respondents were asked to assess the motivations if each country were to increase its presence in the South China Sea (on a seven point scale ranging from 1 (“for purely defensive reasons”) to 7 (“for purely offensive reasons”)).

---

9For other recent survey experiments in China, see Huang (2015); Bell and Quek (2018); Quek and Johnston (2017/18); Weiss and Dafoe (2019).
10For more on how our study addressed social desirability bias, see our discussion in Appendix §1.
11Like with most online surveys in China, the sample overrepresents younger, college-educated respondents in urban areas. See Appendix §1.
12For more details on the survey instrument and translation procedures, see Appendix §2.
Figure 1: Experimental designs

Introduction
Describe tensions in South China Sea

Experimental modules

Perspective-taking module
Manipulate behavior of the other side
- Other side escalates
- Other side de-escalates

Measure DV: Policy Choice
- Should your side escalate or de-escalate? Why?

Attribution module
What inferences would you draw if:
- Other side escalates
- Our side escalates

Order manipulation

Dispositional questionnaire

Demographics
Foreign policy orientations
Empathy measure (Davis 1983)
Image theory measures (cf Herrmann and Fischerkeller 1995)
k-level reasoning (Stahl and Wilson 1995)
The purpose of this module is to solicit attributions for the other side’s behavior (Canetti, Gubler and Zeitzoff, 2021), testing for the asymmetrical attribution pattern identified as a key ingredient of the security dilemma: to what extent do citizens of each country tend to view the other side’s behavior as more offensively-motivated than one’s own?

The perspective taking experiment has a more elaborate design, which we call a *parallel dyadic experimental design*. In it, we manipulated the behavior of the other side: thus, respondents in the United States were presented with actions by the Chinese, while respondents in China were presented with actions by the United States. This parallel and interactive structure is one of the innovations of our methodological approach. Although there have been a number of important survey experiments in IR fielded in multiple countries (e.g. Tomz and Weeks, 2013), they tend to be designed to be studied independently, rather than exploring how the dynamics in one country interact with those in another. Indeed, it is striking that whereas the original audience cost model is dyadic and incorporates strategic interactions between both sides, most of the public opinion research testing its microfoundations, like most survey experiment work more generally, has focused solely on results within a single country. In this sense, one of our efforts here is to bridge the divide between the crisis bargaining literature and experimental studies of public opinion, and to follow the exhortations of Johnston (2012), Bell and Quek (2018) and others in directly testing our theories of IR in an East Asian context.

In the perspective taking experiment, respondents were randomly presented with one of two different actions carried out by the other side.\(^{13}\) Half the sample considered a scenario in which the other side decided to *decrease* naval deployments in strategic maritime zones in the South China Sea, whereas the other half of the sample considered a scenario in which the other side decided to *increase* deployments. To strengthen the treatment, participants were also given a writing task asking them to write out what conclusions their country would draw from that action.\(^{14}\)

Participants in the control condition were then asked to indicate the extent to which they wanted their country’s military activities to increase or decrease, administered using a branching item used to produce a seven-point scale ranging from 1 (decrease a lot) to 7 (increase a lot). However, prior to being administered this item, participants in the perspective taking treatment condition were

---

\(^{13}\) The samples were well balanced across treatments, as shown in Appendix §4.2.

\(^{14}\) See Epley, Caruso and Bazerman (2006) for a similar perspective taking treatment. Utilizing writing tasks to strengthen treatment dosages is a standard technique in the experimental literature: see e.g. Albertson and Gadarian 2016.
induced to engage in perspective taking. Participants in the treatment condition were given an additional writing task, asking them to write what conclusions *the other side would draw* if their own side were to escalate: that is, Chinese participants in the treatment condition were asked what conclusions the Americans would draw if China were to escalate, while American participants in the treatment condition were asked what conclusions the Chinese would draw if the United States were to escalate. Participants in the treatment condition were then administered the same dependent variable administered to participants in the control. Participants in both conditions were also given an open-ended response asking them to explain their answer. Whereas the purpose of the attribution module is to test for attribution asymmetries, the purpose of the perspective taking experiment is to see whether perspective taking can mitigate escalatory dynamics.

Importantly, our experimental design is both parallel and asymmetric: although the structure of the experiments are identical across both samples, the knowledge concepts activated by perspective taking in each side are not, since as its name suggests, the South China Sea is closer to China than to the United States, and is of significantly greater salience for the former than the latter. The research design thus deliberately captures one of the key challenges underlying US security concerns in East Asia, in which many of the potential disputes the US worries about take place in another great power’s backyard; this is the precise strategic context commentators focus on when discussing potential security dilemmas in US-Chinese relations. In this sense, the experimental design has greater internal validity and mundane realism than one about a hypothetical dispute over a fictional territory equidistant from each country’s mainland. We also fielded a follow-up study of our attribution experiment involving a hypothetical territorial dispute taking place off the shores of Hawaii. This follow-up study, which was fielded in the autumn of 2019 on a convenience sample of N=627 American adults recruited through Amazon Mechanical Turk, is described in greater detail below.

Although the study design is somewhat elaborate, it enables us to study five theoretically valuable quantities of interest:

1. By comparing the attributions respondents offer for Chinese and American escalation in the

---

15 The study thus can be thought of as an encouragement design, in that participants were experimentally encouraged to engage in perspective taking, but some individuals may have already been engaging in perspective taking in the control, while others who were administered the treatment may fail to comply with it. We consider the implications for the interpretation of our findings in Appendix §4.1, showing it renders the results reported in the main text a more conservative test.

16 Following Burleigh, Kennedy and Clifford (2018), we drop respondents using Virtual Private Servers (VPS) to mask their location, or whose locations we could not validate as being in the United States.
attribution module, we can measure attribution asymmetries, indicating how much more offensively-motivated citizens perceive an action to be when it is carried out by the other country rather than by their own, a central psychological microfoundation for the security dilemma.

2. By comparing the average policy choices advocated by American and Chinese respondents in each experimental cell in the perspective taking experiment, we can calculate the baseline level of escalatory preferences advocated by citizens in each country, important for understanding leaders’ domestic incentive structures in a potential dispute between the two countries.

3. By estimating the treatment effect of perspective taking on policy choices, we can test whether experimentally inducing participants to take the perspective of the other side causes them to favor more or less escalatory policies in each country, important for testing the validity of palliative theories of perspective taking.

4. By linking the attribution and perspective taking experiments together, we can test whether perspective taking mitigates the escalatory effects of attribution asymmetries — or whether it in fact exacerbates it.

5. By estimating the treatment effect of perspective taking on the explanations respondents give for their policy choices, we can test whether experimentally inducing participants to take the perspective of the other side changes the kinds of justifications they offer for their policy preferences, important for understanding the knowledge structures perspective taking activates.

3 Results

We structure our presentation of the results around the five quantities of interest enumerated above. We begin with the attribution experiment, looking for the central psychological microfoundation of the security dilemma by calculating the extent to which respondents in each country displayed asymmetrical attribution patterns, showing that respondents tend to see actions as significantly more offensively-motivated when carried out by the other side. This attribution asymmetry is especially pronounced in China, and is not simply an artifact of the geographic proximity of the South China Sea, but rather reflects a fundamental asymmetry in each side’s perception of the US-China relationship more generally. Having shown that this pattern exists, we then turn to the perspective taking
experiment, showing how perspective taking encourages reciprocity among American respondents, but opportunistic behavior among Chinese respondents. Rather than perspective taking enabling individuals to better escape security dilemma thinking, it only displays palliative effects among respondents who aren’t prone to attribution asymmetries in the first place. Finally, we obtain similar findings using a series of Structural Topic Models (STMs) to explore how the perspective taking treatment affects the explanations participants offer for their policy choices.

3.1 Attribution module: identifying security dilemma thinking

At its broadest level, the security dilemma is often understood to be a consequence of asymmetric attributions, as each side perceives their own behavior as defensively-motivated, and the other side’s behavior as offensively-motivated (Jervis, 1978). Our attribution module lets us look for individual-level evidence consistent with this theoretical argument.

We therefore asked participants in our study to assess the motivations for a hypothetical escalation in the South China Sea by each side, on a scale ranging from “For purely defensive reasons” to “For purely offensive reasons." Consistent with the logic of security dilemma thinking, in which actors are more likely to perceive their own behavior as defensively motivated than the behavior of others, the results, presented in Figure 2, reveal a significant attribution asymmetry among both Americans (point estimates depicted in solid blue) and Chinese (point estimates depicted in solid red). In each country we find that respondents are much more likely to attribute their country’s increased military activities to defensive motivations and the increased military activities of the other country to offensive motivations. Audiences in both countries believe their side is acting defensively, and the other side is acting offensively. Interestingly, however, the magnitude of the attribution asymmetry is 3.1 times greater in China than the US, with the attribution asymmetry being 2.89 points ($p < 0.001$) among Chinese and 0.94 points ($p < 0.001$) among Americans on the seven point offensive-defensive attribution scale.

There are two interpretations of this differential attribution asymmetry. The first is that Chinese respondents are especially prone to attribution asymmetries in US-China relations more generally. For example, while all great powers have their own narratives of exceptionalism, Chinese political elites routinely emphasizes a self-conception of China as a uniquely peaceful country, more defensively oriented than its western counterparts, with Xi Jinping claiming in speeches that there is “no gene for invasion in Chinese people’s blood” (Johnston, 2018). If Chinese political culture systematically
Figure 2: Respondents display the attribution asymmetry consistent with security dilemma thinking.

Note: Figure 2 displays the average attribution made by Chinese and American respondents for an escalation by the other side ("They increase") and by their own side ("We increase"). If the attribution made for the other side’s behavior is higher on the scale than the attribution made for one’s own behavior, then the results display an "attribution asymmetry", an asymmetry in the motivation attributed to an identical action carried out by both sides. The steeper the slope of the line connecting the two points, the stronger the asymmetry. Thus, the figure shows that while both sides tend to see their own actions as more defensive and the other side’s actions as more offensive, the Chinese public displays a much stronger attribution asymmetry than the American public, which holds even when we relocate the dispute to the shores of Hawaii in a follow-up study, suggesting this larger attribution asymmetry can’t simply be an artifact of geographic proximity.
emphasizes Chinese peacefulness and American negative intentions more generally, one might expect Chinese respondents to display particularly large attribution asymmetries in the context of our experiment. The second is that this asymmetry is due to the geographic proximity of the dispute, which makes American escalation mean something very different in China’s backyard than Chinese escalation does. We therefore tested the effect of geographic proximity by fielding a follow-up study where we replicated the attribution module from our main study, but this time set the potential dispute in America’s backyard: maritime zones off the coast of Hawaii. The study is discussed in detail in the Appendix §4.6.17

To test how much of the difference in attribution asymmetries between the Chinese and Americans is due to the geographic features of the scenario, we compare the magnitude of the attribution asymmetries from the solid and dashed blue point estimates in Figure 2. The results show that when the dispute is relocated to the shores of Hawaii, Americans perceive Chinese escalation as significantly more offensively-motivated ($p < 0.001$) than they did when it was taking place in the South China Sea. However, even though relocating the dispute shrinks the magnitude of the gap between American and Chinese attribution asymmetries, it does not eliminate it, as a comparison between the solid red and dashed blue lines shows: the attribution asymmetry the Chinese provide in the first experiment remains 2.2 times larger than the attribution asymmetry the Americans provide in the second. Since geographic proximity appears to account for only 18% of the gap in attribution asymmetries between the US and Chinese samples in the first experiment,18 this shows that the higher levels of attribution asymmetries we detect in the Chinese sample are not merely an artifact of geographic proximity, and likely also reflect broader differences in how Chinese and American respondents understand US-China relations more generally.

To test this directly, as part of our dispositional questionnaire described in Appendix §3, we build off of work on image theory by including a set of items measuring the images that Americans and Chinese have of one another (Herrmann and Fischerkeller, 1995). As Figure 3 shows, Americans and Chinese understand the US-China relationship in fundamentally different ways. Chinese respondents are significantly more likely to view the United States as harboring threatening intentions than Americans indicate with respect to China ($W = 1385166, p < 0.001$). And, citizens in the two

---

17 Unfortunately, changes in the Chinese political environment following the fielding of our first study meant we were unable to field the Hawaii version of the experiment in China.

18 If $a_{i,j}$ denotes the attribution asymmetry for respondents from country $i$ (either China, or the United States) in a dispute in context $j$ (either the South China Sea, or Hawaii), the proportion of the differential attribution asymmetry we attribute to geographic proximity is calculated as $\frac{a_{c,scs} - a_{u,scs}}{a_{c,scs} - a_{u,haw}}$. 

17
countries also disagree on their relative power ($W = 611282, p < 0.001$): at the time the study was fielded, Chinese respondents perceived China as significantly weaker than the United States, while Americans were more likely to perceive China as having already caught up. Given the centrality of disagreements about relative capabilities in the IR literature (e.g. Kugler and Lemke, 2000; Powell, 2006), the fact that citizens in the two countries disagree about their relative power to such an extent is striking.

Figure 3: Chinese and Americans perceive the US-China relationship very differently

![Image of Other: Negative Intentions](attachment:image.png)

![Image of Other: Equal Capabilities](attachment:image.png)

Note: Chinese and American respondents perceive the US-China relationship very differently more generally. The Chinese perceive the US as harboring significantly more negative intentions than Americans attribute to China, and Americans are more likely to think China has already caught up in power, whereas the Chinese perceive China as far behind.

Crucially, while American and Chinese respondents make systematically different attributions for each country’s behavior in the South China Sea, and display systematically different perceptions of the US-China relationship, there is also considerable heterogeneity of views within each side: when we estimate the intraclass correlation from a set of mixed-effect models on the attributions shown in Figure 2, for example, there is 2.3-2.6 times more variation in attributions within each sample than between them. Table 1 shows we can use this within-sample variation in respondents’ images of the other country to explain this variation in attribution asymmetries as well. In both the United States and China, respondents who perceive the other side as harboring more negative intentions in general also display significantly larger attribution asymmetries in the specific South China Sea
Table 1: Explaining crossnational variation in attribution asymmetry

<table>
<thead>
<tr>
<th></th>
<th>US Sample</th>
<th>China Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Negative intentions</td>
<td>2.659***</td>
<td>1.796***</td>
</tr>
<tr>
<td></td>
<td>(0.256)</td>
<td>(0.274)</td>
</tr>
<tr>
<td>Equal capabilities</td>
<td>-0.798***</td>
<td>-0.610***</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.265)</td>
</tr>
<tr>
<td>Nationalism</td>
<td>0.097</td>
<td>1.079***</td>
</tr>
<tr>
<td></td>
<td>(0.249)</td>
<td>(0.230)</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>2.248***</td>
<td>1.442***</td>
</tr>
<tr>
<td></td>
<td>(0.350)</td>
<td>(0.337)</td>
</tr>
<tr>
<td>Education</td>
<td>0.673***</td>
<td>-1.566***</td>
</tr>
<tr>
<td></td>
<td>(0.249)</td>
<td>(0.419)</td>
</tr>
<tr>
<td>Male</td>
<td>0.011</td>
<td>-0.370***</td>
</tr>
<tr>
<td></td>
<td>(0.124)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.531**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.249)</td>
<td></td>
</tr>
<tr>
<td>Party Member</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.253</td>
<td>-1.696***</td>
</tr>
<tr>
<td></td>
<td>(0.197)</td>
<td>(0.268)</td>
</tr>
<tr>
<td>N</td>
<td>1.491</td>
<td>1.459</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.066</td>
<td>0.107</td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01
dispute. We also find that respondents in both countries who see the two countries as closer together in power tend to perceive smaller attribution asymmetries, especially in China, where this effect is particularly large. In other words, the Chinese are especially prone to the attributional asymmetry at the heart of psychological models of the security dilemma not just because the context where a potential dispute would most likely take place is located within China’s own backyard, but also because of the more threatening image they have of the US in general, which acts as a prism they use to interpret American behavior in specific instances, consistent with how images work in foreign policy more generally (Herrmann and Fischerkeller, 1995).

In aggregate, these findings suggest that in both countries, public opinion during a dispute in the region would more likely be an accelerator of a potential spiral model than an inhibitor against it, which is consistent with the microfoundations of the security dilemma. Both Americans and Chinese display attribution asymmetries consistent with the logic of the security dilemma that could fan the flames of escalation. Our interest in the next experiment, then, is whether, consistent with prevailing wisdom in IR, perspective taking can mitigate these effects.

3.2 Perspective taking experiment

To examine the effects of perspective taking on escalation preferences, the vignette in the perspective taking experiment focuses respondents’ attention on the South China Sea and the security policies of the US and China. Our primary dependent variable is respondents’ preferred policy with regard to military escalation in the South China Sea. Higher values indicate respondents wanted their government to increase its military activities, while lower values indicate a preference for decreasing military activities. Since the dependent variable follows both the escalation (de-escalation) treatment and the perspective taking treatment in our experiment, Figure 4 presents the average policy preferences for respondents in each country, conditioning on treatment assignment.

As Figure 4 shows, and consistent with the results of the attribution module, the baseline preference of military escalation significantly differs between the Chinese and American publics. Chinese respondents are significantly more likely to support increasing their military activities in the South China Sea, with their average preferred policy scoring 1.07 ($p < 0.001$) points higher than the US respondents on a seven-point scale. Regardless of treatment condition, the Chinese public prefers substantially higher levels of escalation in the South China Sea than their American counterparts, which is consistent with the conflict’s proximity and perceived threat for the Chinese.
Figure 4: Effects of treatments on policy preferences

Note: Higher values indicate more escalatory policy preferences. The figure shows that respondents in both countries are more supportive of escalation when the other side does the same, but that the Chinese public are consistently more supportive of escalation than the Americans are. Interestingly, the effect of the perspective taking (depicted here with p-values from t-tests) is highly contingent. Consistent with palliative accounts of perspective taking in IR, perspective taking makes Americans less escalatory when China does the same, facilitating positive reciprocity. However, for Chinese respondents the effect of perspective taking is inverted: when the United States de-escalates, the Chinese respond to the perspective taking treatment by favoring greater escalation, while when the United States escalates, the Chinese respond to the perspective taking treatment by favoring greater de-escalation.
We next turn to the effects of the escalation treatment by comparing the left and right panels of Figure 4. Notably, the results show the escalation patterns consistent with the spiral model that fuels the security dilemma. In both samples, respondents prefer higher levels of escalation by their own government after the other country escalates their military activities, with the average increase for the US and China being 0.50 and 0.59 respectively (both significant at $p < 0.001$). As Jervis (1976, 84-92) notes, this reciprocal pattern of escalation is more consistent with the spiral model than the deterrence model, in that escalation is prompting further escalation, rather than increasing the likelihood of the other side wanting to back down. Combined with the attribution asymmetry results, we find that, on the whole, American and Chinese respondents are viewing their countries’ military activities in our scenarios in ways consistent with the microfoundations of the security dilemma.

Most important are the effects of our perspective taking treatment, which displays divergent effects on policy preferences in the US and China. In the US, the perspective taking treatment causes participants to advocate for de-escalation when the other side is observed decreasing their military activities ($-0.27$, $p < 0.018$). However, we find that the perspective taking treatment has the opposite effect among the Chinese public. After engaging in perspective taking and observing the US de-escalate its military activities, Chinese respondents prefer significantly higher levels of military activities ($0.23$, $p < 0.037$). Rather than serving as a panacea, perspective taking instead encourages Chinese respondents to strategically seek to take advantage of their adversary’s de-escalation by moving to gain a decisive advantage in the conflict. Interestingly, the perspective taking treatment also has the effect of decreasing the hawkishness of the Chinese public’s preferred policy when the US has already escalated its military activities ($-0.23$, $p < 0.017$). The fact that perspective taking induces inverse reciprocity among Chinese respondents suggests that it activates deterrence schema (Powell, 1987), in which it makes the Chinese respondents more likely to be deterred from military escalation when the US escalates, but also more likely to escalate when the US shows weakness; American behavior becomes seen through the prism of costly signals of resolve, rather than trust (Kydd, 2005).

---

19 Given potential non-compliance, in Appendix §4.1 we combine sensitivity analyses with an instrumental variable approach to estimate causal average complier effects, showing how the results presented in the main text are likely underestimating the effect of the treatment.

20 Perspective-taking involves cognitive exertion, which means perspective-taking inherently involves rumination, which is a causal mechanism through which perspective-taking affects downstream attitudes or behavior. This point is worth emphasizing, since it suggests that divorcing perspective-taking from rumination entirely would be a form of post-treatment bias (King and Zeng, 2007). However, respondents in both the control and treatment may engage...
To better understand the divergent effects of perspective taking across the US and Chinese samples—especially in the de-escalation condition, where the backfire effect in the Chinese sample is especially striking given the palliative effects assumptions that pervade IR scholarship—we turn to psychological research on the dark side of perspective taking, which suggests that negative effects are especially likely when actors perceive threats to their identity, and when actors perceive threats to their goals or interests. For each of these potential mechanisms, we adopt a two-pronged empirical strategy, in which we first test whether the US and Chinese samples differ from one another along a given dimension, and then test whether variation in that dimension can help explain within-sample variation in the effect of the perspective taking treatment.

While there are many mechanisms through which identity threats operate in IR (e.g. Mitzen, 2006; Rousseau, 2006), the most important is national attachment, reflecting the extent to which respondents identify with their country (Powers, 2021). More nationalist individuals are not only more likely to feel like being a member of their national ingroup is an important part of their identity, but also have a starker perception of shared fate. As Herrmann (2017, S62) writes, “The more someone attaches his or her own identity to the nation, the more they will feel the possible threats to the nation” as a result. To measure respondents’ levels of nationalism, we measured national attachment in both our American and Chinese samples, described in Appendix §3. We find that Chinese respondents are significantly more nationalist than American respondents are (t = 12.99, p < 0.001), as shown in Figure 5(a) below, and Table 5 in Appendix §4.3 shows that more nationalist respondents are more supportive of escalation.

Yet in order for nationalism to explain the divergent effects of perspective taking between the two samples, it would need to moderate the effects of the perspective taking treatment. However, we find weak evidence in support of nationalism’s moderation effect, as shown in Figure 5(b): although three of the effects shown here have the right sign—nationalist respondents engaged in perspective taking are more likely to want to escalate, and less nationalist respondents engaged in perspective taking are less likely to do so—the interaction effects aren’t statistically significant. Even though the Chinese respondents appear more nationalist than the American ones are, we do not have strong evidence that identity threat can explain the differential effects of perspective taking between China in varying levels of rumination, and so we assessed the extent to which rumination itself contributes to the effect of perspective taking. To do so, we use measures of response latency, under the assumption that respondents who spent more time engaging with the scenario engaged in more rumination than respondents with shorter response latencies. As shown in Appendix §4.5 we do not find that response latency explains the effects of perspective taking, suggesting that the perspective taking treatment results in more than simply a rumination effect.
Figure 5: Testing the identity threat mechanism

Panel a shows that the Chinese sample (in red) is more nationalist than the American sample (in blue). Panel b depicts the marginal effect of the perspective taking treatment at different levels of nationalism in all four treatment cells. While 3 of the 4 interaction effects have the right sign (in which perspective taking increases the likelihood of escalation among more nationalist respondents), the slopes are modest and the interaction effects are not statistically significant. Effects estimated using a kernel estimation procedure with k-fold least-squares cross-validation using the interflex package in R to avoid making functional form assumptions, and control for a range of demographic characteristics.

The second context where perspective taking can backfire is when actors vary in their perceptions of threats to their goals. In situations of positive goal interdependence, perspective taking can promote prosocial behavior towards the outgroup, but in situations of negative goal interdependence, perspective taking can heighten awareness of conflicts of interest (Okimoto and Wenzel, 2011; Mooijman and Stern, 2016), especially when the perceived capabilities of the other side are relatively high. As discussed earlier and shown in the left-hand side of Figure 3, Chinese respondents perceive greater threats to their interests from the United States than Americans perceive from China ($W = 1385166, p < 0.001$). The important question is how these perceptions of negative intentions moderate the impact of perspective taking.

Figure 6 shows that perceived negative goal interdependence mitigates the palliative effects of perspective taking in both the US and Chinese samples: when Americans who attribute negative intentions to China more generally are asked to engage in perspective taking, they’re less likely to reciprocate Chinese de-escalation ($p < 0.078$), while it’s only those Chinese who attribute more
Figure 3 showed that the Chinese sample perceives the US to harbor more negative intentions in general than the American sample perceives about China. Figure 6 depicts the marginal effect of the perspective taking treatment at different levels of perceived negative goal interdependence in each sample. Chinese respondents who perceive the United States as harboring more negative intentions are much less likely to back down in response to American escalation, while American respondents who perceive negative goal interdependence from the Chinese are less likely to engage in positive reciprocity when the Chinese de-escalate. Effects estimated using a kernel estimation procedure with k-fold least-squares cross-validation using the interflex package in R to avoid making functional form assumptions, and control for a range of demographic characteristics.

We therefore find stronger evidence in favor of the negative goal interdependence mechanism than we do for identity threat.

Finally, we can exploit the fact that respondents completed both the attribution module and the perspective-taking module to tie the two experiments together. The attribution module shows that both American and Chinese respondents are prone to attribution asymmetries, but also that its magnitude varies substantially across respondents in both countries. The perspective taking experiment shows that perspective taking can sometimes have a dark side in foreign policy crises, especially for actors who perceive threats to their identities and goals. If the conventional wisdom

---

21 It is possible that respondents who view the US and China as having an inherently conflictual relationship (those high in perceived negative goal interdependence), view the military interaction between the countries as one focused on deterrence. If so, a costly escalation of force may function as a costly signal of resolve (Fearon, 1994). What is notable in our results is that at high levels of perceived negative goal interdependence, shown on the right sides of each panel in Figure 6, perspective taking does not have a significant effect on preferred levels of escalation. This suggests that perspective taking is not helpful in mitigating preferences for escalation amongst those who already believe the countries have a significant conflict of interest, but it also suggests that perspective taking is not harmful at these levels either.
Figure 7: Security dilemma thinking mitigates the palliative effects of perspective taking

Figure 7 depicts the marginal effect of the perspective taking treatment at different levels of security dilemma thinking. Chinese respondents who displayed the largest attribution asymmetries in the attribution experiment are much more likely to respond to perspective taking in the perspective taking experiment by wanting to escalate when the United States backs down, while American respondents who displayed the largest attribution asymmetries in the attribution experiment are much less likely to respond to perspective taking by wanting to de-escalate when China backs down. Effects estimated using a kernel estimation procedure with k-fold least-squares cross-validation using the *interflex* package in R to avoid making functional form assumptions, and control for a range of demographic characteristics.

about the palliative effects of perspective taking is correct, perspective taking should mitigate the effects of attribution asymmetries, allowing individuals to break out of the security dilemma. Yet as Figure 7 shows, we find no evidence in support of this conventional wisdom. Rather than finding that perspective taking mitigates the effects of attribution asymmetries, Figure 7 shows that security dilemma thinking counteracts the palliative effects of perspective taking: in both the Chinese and US samples in the de-escalate condition, perspective taking only has palliative effects among respondents who are low in attribution asymmetries ($p < 0.004$ for China, $p < 0.007$ for US). High attribution asymmetry respondents are much more likely to respond to perspective taking by wanting to escalate: the stronger the double standard you hold the other country to for its behavior, the less effective perspective taking is, perhaps because perspective taking works not by transcending our knowledge structures about the target, but by further activating them. These findings thus suggest an irony in advocating for perspective taking as an all-purpose solution to attribution asymmetries, since it appears to only have these palliative effects among people who are less prone to negative attribution asymmetries in the first place.

---

22We operationalize low and high here based upon whether respondents display a positive attribution asymmetry or not.
3.2.1 Effects of perspective taking on policy reasoning

The above analysis shows that perspective taking does not necessarily have palliative effects: perspective taking induces reciprocity among American respondents — who are more likely to de-escalate if the Chinese do the same — but not among our Chinese respondents, who seek to project strength when the US de-escalates, but de-escalate in the face of American strength. Our previous analysis suggested that one reason why perspective taking produces different results for each sample is because it activates different knowledge structures in each group, with Chinese respondents perceiving greater threat from the US more generally than Americans perceive from the Chinese.

To further explore the content of these knowledge structures, we turn to the explanations respondents gave for their preferred policy choice. We asked all respondents to explain their preferred policy choice in a free response format. We analyze these open-ended responses using a structural topic model (STM), semiautomated text analysis models that allow us to measure the impact of our perspective taking treatment on the language used by respondents. Thus, whereas the previous analysis asks how experimentally inducing perspective taking affects what respondents want their countries to do, this analysis turns to discourse and asks how perspective taking affects why respondents want their countries to behave in a particular way. An advantage of STMs are that they are unsupervised, which means they “discover” topics in the text, rather than assuming their existence, which limits the potential for the researcher’s prior expectations to influence the topics that are identified. Most importantly for our analysis, unlike more traditional models used in text analysis, such as Latent Dirichlet Allocation (LDA), STMs allow us to leverage information about respondents (in our case, the treatment group to which they were randomly assigned) when structuring the topics, rather than assuming that the topics and topical prevalence are constant across all respondents.

We estimate a series of STMs on our respondents’ open-ended explanations for their policy choice, using the models to identify topics within the responses and, most importantly, to measure whether those topics systematically differ in prevalence across the perspective taking treatment and control groups.²³ Due to space constraints, we focus our analysis on the most theoretically interesting part of the experiment highlighted above: the de-escalate condition, where perspective taking has divergent effects in both samples, leading to reciprocal deescalation in the American sample, but

²³For our US respondents, we pre-process the text using the tm package in R, whereas for our Chinese respondents, we use Jieba, a Python-based Chinese word segmentation module. Once the models have been estimated, we then interpret the English-language versions of the topics, as translated by three native Chinese speakers to ensure intercoder reliability.
For the American sample in Figure 8, we find that when China de-escalates, the perspective taking treatment results in greater usage of language emphasizing reciprocity. Consistent with the positive goal interdependence perceived by American respondents, respondents are much more likely to invoke language embracing reciprocal de-escalation, and do not view the Chinese presence as particularly threatening, with one respondent writing that the US should “respond positively to China’s willingness to compromise” and another noting that “we should move out in the same fashion as the Chinese (meaning ship for ship).” Echoing the analysis above, these findings suggest that the perspective taking treatment makes it more likely that American respondents interpret China’s decrease in military activities as signaling willingness to compromise based on compatible goals, which leads them to explain their policy preferences in terms of reciprocal efforts to de-escalate the situation.

Figure 9 displays the STM results for our Chinese sample. It shows two different ways that the perspective taking treatment significantly affects Chinese participants’ policy explanations.24

---

24In addition, a third topic also significantly differs in topical prevalence, but it is not semantically coherent.
Results from structural topic models, China sample. The top-left panel displays the effect of the perspective taking treatment on the free responses provided by respondents justifying their policy recommendation; the effect estimates are measures of changes in topical prevalence. The remaining panels display representative responses for each topic, translated into English.
First, consistent with Chinese respondents’ perception of negative goal interdependence and threats to Chinese interests, it increases their emphasis on standing firm and defending China’s interests in the region. One respondent wrote “Increase military exercises. Build aircraft carriers as soon as possible, build ten or eight of them”, while another noted that “Only by [escalating] can China have deterrent power in the world. China advocates peace but will not compromise." The escalation Chinese respondents prefer when responding to a US de-escalation in the perspective taking condition reflects this logic, consistent with perspective taking activating deterrence mindsets among Chinese respondents. Second, asking participants to assess the conclusions the United States would draw leads them to be less likely to emphasize the importance of showing moderation: participants in the control condition were significantly more likely to make comments urging both parties to step back (offering responses like “both sides should take a step back, and then they will see the vastness of the sea and the sky.”)

In sum, perspective taking affects both what respondents want their countries to do, and why. Against the palliative effects hypothesis, however, its effects are not always positive, but contingent on perceived threats to actors’ interests. It makes American respondents, who perceive relatively little threat to their interests in the dispute, place a heavier emphasis on reciprocity. In contrast, it makes Chinese respondents, for whom a dispute in the South China Sea activates knowledge structures relating to threat, more interested in standing firm when the US shows signs of weakness.

4 Conclusion

The rise of China has led to an outgrowth of interest in security dilemmas, and a surge of scholarship exploring the ways that tensions between the rising power and the United States can be mitigated. Given continued tensions in the South China Sea, and concerns about popular nationalism in both the American and Chinese publics, understanding how the perceptions held on both sides are formed — and how they are shaped by perspective taking — is a useful endeavor for both scholars and policymakers alike. Our findings are mixed news for psychological microfoundations of the security dilemma, finding strong evidence in favor of the central psychological driver of the security dilemma, but mixed evidence about a potential solution to it.

In our first set of cross-national experiments, we find micro-level evidence consistent with the logic of the security dilemma in both China and the United States: both Chinese and American
respondents perceive the same action as defensively motivated when carried out by their own side, and offensively motivated when carried out by the other. We also show that the magnitude of this attribution asymmetry is partly a function of individuals’ understandings of the US-China relationship more generally, which differ quite dramatically between the two countries.

These findings make a number of important contributions to the study of international security, political psychology, and public opinion in US-China relations. First, although political psychologists have long argued the security dilemma has psychological microfoundations, this is the first study to identify this attribution asymmetry experimentally, which we find evidence of in national samples in two different great powers. Security dilemma thinking is therefore not merely the preserve of elite decision-makers, but is also detectable in the public writ-large. Second, our findings shed new light on public opinion dynamics in the region, a crucial subject in an era when policymakers are increasingly concerned about the extent to which the mass public in both countries can function as an accelerant rather than an inhibitor of great power conflict. Against an older tradition in IR assuming that public opinion has inherently pacifying effects on foreign policy, we show here that if a crisis were to break out in the region, there would be a reservoir of support for escalation on the Chinese side. Yet Americans aren’t immune to attribution asymmetries either: Americans perceive US escalation to be defensively motivated no matter where it takes place, but perceive the same acts by another country to be offensively motivated even when it takes place in that other country’s backyard. Third, the differential perceptions of this US-China relationship we report here have important implications for American foreign policy and debates about American hegemony. IR scholars and pundits in the US often talk about the United States as a uniquely benevolent hegemon, arguing that the fundamentally liberal character of the US means that American hegemony shouldn’t be seen as threatening by others (e.g. Kagan, 1998). Yet as both Figure 3 and the open-ended responses show, our Chinese respondents see the US as wanting to dominate, rather than to lead. Our results are therefore consistent with a dissenting realist tradition arguing that the accumulation of power in the hands of others can be threatening in and of itself: the stronger the Chinese perceive the United States to be, they more they assume that US escalation is offensively motivated.

In our second set of cross-national experiments, our findings add a cautionary note to the pal-liative effects of perspective taking hypothesis often adopted by IR scholars as a potential solution to the attribution asymmetry identified above. IR scholars have tended to think about perspective taking’s effects in relatively liberal ways, consistent with a broader ethos in IR theory that presumes
the existence of a natural harmony of interests, such that conflict is merely the result of misperceptions or bargaining failures (Kertzer, 2016, 41): if only actors are given more information, or decision-makers’ excesses can be constrained by liberal-minded publics, cooperation will be more likely to occur. Our findings push back against both claims. Perspective taking may reveal the existence of complimentary interests, but it can backfire and highlight conflicts of interest as well, particularly when actors perceive threats to their interests. Perspective taking makes Americans – who perceive less threat – more likely to recognize conciliatory signals and respond with reciprocal policies of de-escalation, but Chinese respondents have a very different reaction, causing them to seek to project strength and think less about the regional ramifications when the US de-escalates. Moreover, rather than perspective taking causing individuals to be better able to avoid the security dilemma, we find that perspective taking only has palliative effects among individuals not prone to attribution asymmetries in the first place.

In demonstrating that perspective taking in IR is not an unvarnished good, these findings suggest several avenues for future research. First, our focus here was on cognitive forms of perspective taking, rather than affective forms of perspective taking that emphasize feeling what others are feeling. Future work should test the differential effects between the two. Second, we largely side-step here the question of empathic accuracy: our analysis focuses on what happens when we encourage Chinese and Americans to step into each other’s shoes, rather than assess how competently they do so. Future scholarship can benefit from investigating the accuracy of these second-order beliefs directly (Mildenberger and Tingley, 2019). Third, it is worth noting that our perspective taking treatment was deliberately content-free, seeking to induce perspective taking rather than priming participants to attribute particular beliefs to the other side. While consistent with other studies of perspective taking in psychology (e.g. Epley, Caruso and Bazerman, 2006), the study nonetheless differs from the content-specific perspective taking interventions sometimes used in political science studies seeking to reduce intergroup prejudice. Other work can benefit from employing alternate forms of perspective taking treatments, determining what proportion of the effects come from perspective taking itself rather than the informational primes. In this manner, by showing that perspective taking is a broader construct — and thus has more complex effects — than many IR scholars assume, this study points to a broader research agenda on the dynamics of perspective taking in international affairs.


Weiss, Jessica Chen. 2013. “Authoritarian Signaling, Mass Audiences, and Nationalist Protest in
Oxford: Oxford University Press.
University Press.
White, Ralph K. 1986. “Empathy as an intelligence tool.” International Journal of Intelligence and
CounterIntelligence 1(1):57–75.
Yarhi-Milo, Keren. 2014. Knowing the Adversary: Leaders, Intelligence, and Assessment of Inten-
Perspective Taking and Security Dilemma Thinking:
Supplementary Appendix

Joshua D. Kertzer, Ryan Brutger, and Kai Quek
October 12, 2023

Contents

1 Note on survey partner 2
   Table 1: Sample characteristics: Chinese sample . . . . . . . . . . . . . . . . . . . . 3
   Table 2: Sample characteristics: US sample . . . . . . . . . . . . . . . . . . . . . . 3
   Figure 1: Young, urban, and educated Chinese overrepresented in Chinese sample . . 4

2 Translation protocol for survey instrument 5

3 Survey instrument 5

4 Supplementary analysis 9
   4.1 Causal average complier effects ................................................. 9
      Table 3: Causal average complier effect (CACE) estimates, US sample . . . . . . 9
      Table 4: Causal average complier effect (CACE) estimates, Chinese sample . . 9
   4.2 Balance tests ................................................................. 11
      Figure 2: Balance tests: China experiment ..................................... 11
   4.3 Exploiting natural variation in perspective-taking motivation .................... 11
      Figure 3: Balance tests: US experiment ........................................ 12
      Table 5: The dispositional perspective-taking results offer little support for the pal-
                  liative effects of perspective-taking hypothesis ......................... 13
   4.4 Structural topic model results in the escalation conditions ....................... 14
      Figure 4: Structural topic model results: China (United States escalates condition) 14
   4.5 Rumination Effects .................................................................. 15
   4.6 Follow-up attribution experiment: Hawaii ........................................... 15
1 Note on survey partner

Political surveys are sensitive in China. This is especially the case for a national-level survey fielded across all provinces and capital municipalities of China; as Table 1 shows, the resulting sample skews closely to demographic targets (as does the US sample - see Table 2). Given the political conditions, our partner company would only agree to field our experiment on the condition of strict confidentiality. We persuaded our partner to make an exception for the review process; however, no identifying information can be revealed to others beyond this exception as the leakage may impact the company’s interests in China.

While political/social desirability bias can exist in any survey sample, the concern may be greater in China due to the political environment. To manage this concern, we opted for an Internet survey instead of a face-to-face or telephone survey. Researchers have shown that anonymous online surveys reduce desirability bias (Chang and Krosnick, 2010). Indeed, we found that respondents seem to be quite willing to disapprove of the government when it does something that goes against their individual preferences. For example, in a 2015 study, one of the co-authors found that the majority of respondents did not approve of a Chinese leader who had threatened to use force in a territorial dispute but subsequently did not. Of course, even if desirability bias exists in China, it does not make the results less useful. The results remain relevant because the same desirability bias also operates in the real world.

The US sample skews closely to national demographic targets (see Table 2). As Table 1 and Figure 1 show, while our online Chinese sample adheres fairly closely to census targets in terms of gender, region, ethnicity and marital status, as is the case with most online samples in China, it over-represents the young, urban and educated – respondents over 65 and agricultural workers tend to be particularly underrepresented. Two points are worth noting here. First, as Johnston (2017) argues, the Chinese leadership is particularly attentive to expressions of popular nationalism in large urban centers – the same centers overrepresented in our results. If Chinese political elites are more likely to be monitoring the views of younger, urban and more internet-savvy citizens, this makes the patterns in our results more politically important.

Second, these compositional features of our sample should be the most consequential for the sample-level descriptive statistics we report – such as when comparing the distributions of nationalism and perceptions of negative intentions between the US and Chinese sample. Yet the existing literature has reported mixed findings about the relationships between the demographic characteristics overrepresented in our sample and nationalism or images of the United States. For example, for age, Tang and Darr (2012, 820) find that “the socialist and Cultural Revolution generations are more nationalistic than the reform and the post-reform generations” and Wang et al. (2022, 245) find that younger Chinese “espouse more favorable attitudes toward" the United States, while Weiss (2019, 684-85) found that younger Chinese “were more hawkish than their elders", and Fang, Li and Liu (2022, 40-41) find that younger Chinese tend to hold more negative views of the United States. For education, Wang et al. (2022, 245) find that “it is the least-educated respondents who hold the least-favorable attitudes towards the U.S", while Fang, Li and Liu (2022, 41) find that more educated respondents have a more negative view of the China-US relationship, and Tang and Darr (2012, 819) find a curvilinear relationship between age and nationalism, where the least and most educated are less nationalistic than individuals with moderate levels of education. Similarly, Tang and Darr (2012, 819) report that “urban residents have the lowest level of nationalism, lower than both rural residents and rural migrants", while Johnston (2017, 16) find that levels of nationalism in Beijing are similar to those expressed in nationwide studies. In sum, then, we do not have strong theoretical priors that lead us to believe our sample is skewing our results in a particular direction.
Table 1: Sample characteristics: Chinese sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Full Population</th>
<th>Age ≥ 18</th>
<th>2016 sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>51.2%</td>
<td>50.5%</td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern</td>
<td>12.7%</td>
<td>13.1%</td>
<td></td>
</tr>
<tr>
<td>Northeastern</td>
<td>8.8%</td>
<td>9.7%</td>
<td></td>
</tr>
<tr>
<td>Eastern</td>
<td>30.2%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>Central &amp; Southern</td>
<td>27%</td>
<td>25.9%</td>
<td></td>
</tr>
<tr>
<td>Southwestern</td>
<td>14.2%</td>
<td>15.9%</td>
<td></td>
</tr>
<tr>
<td>Northwestern</td>
<td>7.2%</td>
<td>7.3%</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 19</td>
<td>24.1%</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>20-29</td>
<td>17.1%</td>
<td>21.7%</td>
<td>25.6%</td>
</tr>
<tr>
<td>30-39</td>
<td>16.1%</td>
<td>20.4%</td>
<td>24.1%</td>
</tr>
<tr>
<td>40-49</td>
<td>16.3%</td>
<td>21.9%</td>
<td>18.2%</td>
</tr>
<tr>
<td>≥ 50</td>
<td>25.3%</td>
<td>32%</td>
<td>30%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Han</td>
<td>91.6%</td>
<td>92.3%</td>
</tr>
</tbody>
</table>

Table 2: Sample characteristics: US sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample</th>
<th>Population</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>0.45</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.55</td>
<td>0.51</td>
</tr>
<tr>
<td>Age</td>
<td>18-24</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>25-44</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>45-64</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>≥ 65</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Education</td>
<td>High school or less</td>
<td>0.34</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Some college/university</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>College/university</td>
<td>0.37</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Graduate/professional school</td>
<td>0.17</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Figure 1: Young, urban, and educated Chinese overrepresented in Chinese sample

(a) Age

(b) Gender

(c) Region

(d) Education

(e) Ethnicity

(f) Sector

(g) Marital Status

(h) Urban-Rural

Sample

Census (2010)

Sample
2 Translation protocol for survey instrument

The survey instrument was first written in the English language. For our Chinese respondents, the instrument was translated into Chinese and pre-tested in multiple stages to ensure accuracy. First, a native Chinese speaker (C1) translated the English instrument into Chinese. Next, one of the co-authors (A1) vetted the translated text and produced the first pre-test version. Both C1 and A1 are effectively bilingual. The focus of the translation was accuracy of meaning instead of a literal word-to-word translation, which can cause meaning distortions when the translated words combine into sentence form. Subsequently, this survey version is tested separately with four other Chinese speakers (C2, C3, C4, C5). Their feedback was used by A1 to produce the second pre-test version. This version is pre-tested with a small sample of Chinese natives (NS) selected from each of the different regions of China to check if the questions were clear and natural to them. Unclear or awkward wordings (few were detected at this stage) were revised in the final translated version. Sample NS was monolingual and not aware of the study design. Finally, the Chinese-language survey was also pre-tested by four additional Chinese language speakers, one native and three bilingual, to flag any problematic or confusing language.

3 Survey instrument

Note: the instrumentation below is for the version of the study fielded in China, with the text also translated into English. In the United States version of the study, the instrumentation is identical, except all references to China are instead references to the United States, and vice versa (e.g. in the American version, respondents are told that China has decided to increase/decrease naval deployments, whereas in the Chinese version, respondents are told that the United States has decided to increase/decrease naval deployments.)

Perspective-Taking Experiment

Introduction

The following questions are about China’s relations with other countries around the world.

Recently there has been much attention over tensions in the South China Sea. Multiple countries in the region have claimed rights to disputed international waters, which are home to a wealth of natural resources, fisheries, and trade routes — all of which are at stake in the increasingly frequent diplomatic standoffs. China is concerned about the United States’ assertiveness in the region.

Writing Task I

Suppose the U.S. has decided to [increase / decrease] naval deployments in strategic maritime zones in the South China Sea. In your view, what conclusion would China draw from the U.S.’s action? [Writing task I: Open-ended response]

Perspective-taking treatment

• 海南近来局势紧张。
中国国内在讨论，是否应当改变中国在南海的军事存在。您认为，若中国决定增加或减少其在南海的军事行动，美国会从中得出什么样的结论？

**Background:**
- There has recently been tension in the South China Sea.
- Multiple countries in the region have claimed rights to disputed waters. China is concerned about the U.S.'s assertiveness in the region.
- Suppose the U.S. has decided to [increase / decrease] naval deployments in strategic maritime zones in the South China Sea.

There is a debate in China about whether China should change their military presence in the South China Sea.

If China were to increase its military activities in the South China Sea, in your view, what conclusion would the U.S. draw from China’s action? [Writing task II: Open-ended response]

**Policy choice**

In this scenario, do you think that China’s military activities around the South China Sea should increase, decrease, or be kept about the same? [Increase / decrease / kept about the same]

**If increase**

中国的军事行动应当增加很多，还是增加一些？

[增加很多 / 增加一些]

Should China’s military activities increase by a lot, or only somewhat? [Increase / decrease / kept about the same]

[If decrease]

中国的军事行动应当减少很多，还是减少一些？

[减少很多 / 减少一些]

Should China’s military activities decrease by a lot, or only somewhat? [Increase / decrease / kept about the same]

[If kept about the same]

您是倾向于增加军事行动、倾向于减少军事行动、还是不倾向于任何一方？

[倾向于增加 / 倾向于减少 / 不倾向于任何一方]

Do you lean toward increasing military activities, lean toward decreasing them, or don’t lean either way? [Increase / decrease / kept about the same]

请简要解释您的回答。

Please briefly explain your response.

**Attribution module**

对于下列说法，请选择最接近您观点的回答：
For each of the following statements, please select the answer that most closely matches your own views.

- If China were to increase its military activities in the South China Sea, do you think it would be: [1 (Purely for defensive reasons) – 7 (Purely for offensive reasons)]
- If the U.S. were to increase its military activities in the South China Sea, do you think it would be: [1 (Purely for defensive reasons) – 7 (Purely for offensive reasons)]

For each of the following statements, please select the answer that most closely matches your own views.

- If China were to increase its military activities in the South China Sea, how do you think it would affect the possibility of Sino-American economic cooperation? [1 (More likely to cooperate) – 7 (Less likely to cooperate)]
- If the U.S. were to increase its military activities in the South China Sea, how do you think it would affect the possibility of Sino-American economic cooperation? [1 (More likely to cooperate) – 7 (Less likely to cooperate)]

**Dispositional measure of perspective taking**

Respondents also completed a dispositional questionnaire alongside the usual battery of demographic characteristics. Most notably, the questionnaire included a dispositional measure of the propensity to engage in perspective taking: the perspective-taking sub-scale of the interpersonal reactivity index (IRI) from Davis (1983), a commonly-used cognitive measure of empathy in the psychological literature. Whereas the perspective-taking treatment experimentally encourages half of our participants to engage in perspective-taking, the perspective-taking scale measures perspective-taking as a trait, scoring participants based on the extent to which they agree with the following questions. To avoid downstream effects, these dispositional questions were randomly assigned in an order manipulation, such that some participants answered the questionnaire prior to the scenario, and others after.

- I try to look at everybody’s side of a disagreement before I make a decision.
- I believe that there are two sides to every question and try to look at them both.
- Before criticizing somebody, I try to imagine how I would feel if I were in their place.
- 我在做决定前会尽可能参考每个人的反对意见
- 我认为每个问题都有正反两面，并尽可能通盘考虑两个方面
- 在批评他人之前，我会考虑，如果我处在他人的情况下我会怎么想

**Other dispositional measures**
We measure nationalism using a question that measures respondents’ level of attachment to their country. Specifically, we use an item from the national attachment scale of Herrmann, Isernia and Segatti (2009), which asks:¹

- When someone says something bad about the American people, how strongly do you feel it is as if they said something bad about you?
- 当有人说中国人坏话时，您在多大程度上觉得这与说您本人坏话一样？

We measure respondents’ images of the the other country using two items. Respondents are prompted with “For each of the following statements, please select the answer that most closely matches your own views.” The first item measures perceived negative goal interdependence/threat perception (respondents in the United States were asked about China, while respondents in China were asked about the United States), while the second measures perceptions of relative power capabilities.

- China wants to dominate the world.
- With regard to national power, China has caught up with the United States.

对于下列说法，请选择最接近您观点的回答:

- 美国想要称霸全球
- 以国家实力而言，中国已赶上了美国

Finally, we measure military assertiveness (sometimes referred to as hawkishness) using standard items from Kertzer and Brutger (2016), using three items in the US sample, and two items in the Chinese sample, all of which were measured on five-point Likert scales.

- The best way to ensure peace is through American military strength.
- The use of military force only makes problems worse.
- Going to war is unfortunate, but sometimes the only solution to international problems.

- 中国的军事实力是确保和平的最佳手段
- 战争虽然不幸，但有时候却是解决国际问题的唯一方法

¹The United States survey had a five-point response scale, while the Chinese survey had a four-point response scale. Higher values represent stronger national attachment.
4 Supplementary analysis

4.1 Causal average complier effects

Table 3: Causal average complier effect (CACE) estimates, US sample

<table>
<thead>
<tr>
<th>τ</th>
<th>China de-escalates</th>
<th>China escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CACE 95% CIs</td>
<td>CACE 95% CIs</td>
</tr>
<tr>
<td>0</td>
<td>-0.27 (-0.5, -0.05)</td>
<td>-0.27 (-0.5, -0.05)</td>
</tr>
<tr>
<td>5</td>
<td>-0.35 (-0.64, -0.07)</td>
<td>-0.28 (-0.52, -0.05)</td>
</tr>
<tr>
<td>10</td>
<td>-0.36 (-0.67, -0.07)</td>
<td>-0.25 (-0.5, -0.01)</td>
</tr>
<tr>
<td>15</td>
<td>-0.39 (-0.72, -0.08)</td>
<td>-0.26 (-0.53, 0)</td>
</tr>
<tr>
<td>20</td>
<td>-0.42 (-0.78, -0.08)</td>
<td>-0.28 (-0.56, -0.01)</td>
</tr>
<tr>
<td>25</td>
<td>-0.44 (-0.81, -0.08)</td>
<td>-0.35 (-0.63, -0.06)</td>
</tr>
<tr>
<td>30</td>
<td>-0.47 (-0.87, -0.09)</td>
<td>-0.4 (-0.7, -0.1)</td>
</tr>
<tr>
<td>35</td>
<td>-0.52 (-0.96, -0.1)</td>
<td>-0.42 (-0.74, -0.09)</td>
</tr>
<tr>
<td>40</td>
<td>-0.55 (-1.02, -0.11)</td>
<td>-0.43 (-0.78, -0.09)</td>
</tr>
<tr>
<td>45</td>
<td>-0.61 (-1.13, -0.12)</td>
<td>-0.45 (-0.82, -0.08)</td>
</tr>
<tr>
<td>50</td>
<td>-0.66 (-1.21, -0.13)</td>
<td>-0.43 (-0.82, -0.04)</td>
</tr>
</tbody>
</table>

Table 3 presents CACE estimates, defining compliance at different percentile thresholds (τ) of response length. The first two columns in each panel uses the ratio estimator \( \text{CACE} = \frac{\text{ITT}}{\text{Pr(Complied)}} \) with 95% bootstrapped CIs, while the next two results from a 2SLS model with 95% normal theory CIs. Note that at 0% non-compliance, the CACE estimates replicate the ATEs reported in the main text.

Table 4: Causal average complier effect (CACE) estimates, Chinese sample

<table>
<thead>
<tr>
<th>τ</th>
<th>US de-escalates</th>
<th>US escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CACE 95% CIs</td>
<td>CACE 95% CIs</td>
</tr>
<tr>
<td>0</td>
<td>0.23 (0.01, 0.43)</td>
<td>0.23 (0.01, 0.45)</td>
</tr>
<tr>
<td>5</td>
<td>0.24 (0.01, 0.45)</td>
<td>0.23 (0.01, 0.46)</td>
</tr>
<tr>
<td>10</td>
<td>0.24 (0.01, 0.45)</td>
<td>0.23 (0.01, 0.46)</td>
</tr>
<tr>
<td>15</td>
<td>0.26 (0.01, 0.48)</td>
<td>0.38 (0.14, 0.62)</td>
</tr>
<tr>
<td>20</td>
<td>0.26 (0.01, 0.48)</td>
<td>0.38 (0.14, 0.62)</td>
</tr>
<tr>
<td>25</td>
<td>0.29 (0.01, 0.54)</td>
<td>0.46 (0.19, 0.73)</td>
</tr>
<tr>
<td>30</td>
<td>0.32 (0.01, 0.6)</td>
<td>0.53 (0.24, 0.81)</td>
</tr>
<tr>
<td>35</td>
<td>0.33 (0.01, 0.62)</td>
<td>0.57 (0.28, 0.87)</td>
</tr>
<tr>
<td>40</td>
<td>0.37 (0.02, 0.69)</td>
<td>0.49 (0.17, 0.81)</td>
</tr>
<tr>
<td>45</td>
<td>0.4 (0.02, 0.76)</td>
<td>0.42 (0.08, 0.76)</td>
</tr>
<tr>
<td>50</td>
<td>0.41 (0.02, 0.78)</td>
<td>0.44 (0.09, 0.79)</td>
</tr>
</tbody>
</table>

Table 3 presents CACE estimates, defining compliance at different percentile thresholds (τ) of response length. The first two columns in each panel uses the ratio estimator \( \text{CACE} = \frac{\text{ITT}}{\text{Pr(Complied)}} \) with 95% bootstrapped CIs, while the next two results from a 2SLS model with 95% normal theory CIs. Note that at 0% non-compliance, the CACE estimates replicate the ATEs reported in the main text.

A potential concern when evaluating the effects of perspective-taking is that our experiment is akin to an encouragement design: participants are asked to think through and write out what conclusions the other country would draw from a particular foreign policy behavior. As with any encouragement design, then, it is important to address potential concerns about non-compliance: most notably, some participants in the treatment condition, who are encouraged to engage in perspective-taking may not take the exercise seriously and refuse to engage in it; in a potential outcomes framework, these would be the “never takers.” An examination of the contents of the second writing task suggest clear evidence that some participants did not comply with the treatment, writing responses in the US survey like “Balls”, “Merica” or “no comment.”
To explore the implications of non-compliance for our interpretation of the experimental results, we combine an instrumental variables approach with sensitivity analyses. First, we calculated the causal average complier effects (CACE) using the Angrist, Imbens and Rubin (1996) ratio estimator, in which the CACE is the intention-to-treat (ITT) effect divided by the proportion of compliers in the treatment condition. To avoid making subjective coding decisions about the types of responses that count as complying, we define compliance using a sensitivity analysis approach, defining compliance thresholds at different percentiles of response length to the second writing task completed by participants in the perspective-taking treatment, under the assumption that respondents who wrote longer responses are more likely to have complied with the treatment. We calculate response length in the US survey by simply calculating the number of characters of the free response; for the Chinese survey we first pre-process the free-responses using Jieba, a Python-based Chinese word segmentation module. The ratio estimator results are presented in the first two columns in each panel in Table 3 for the US survey, and in Table 4 for the Chinese survey, with 95% bootstrapped confidence intervals. We also estimate the CACE using a more formal 2-stage least squares (2SLS) approach, instrumenting for compliance with the treatment using random assignment to the perspective-taking treatment (Gerber and Green, 2012); these results are shown in the third and fourth columns in each panel of each table, with 95% normal theory confidence intervals.

Despite the crudeness of our indicators of compliance, we find results largely consistent with our theoretical story. In the US survey, as the response-length cutoff for defining compliance with the perspective-taking treatment increases, the effect estimates for respondents exposed to the China de-escalate treatment becomes larger, meaning those who complied with the treatment were more likely to prefer de-escalatory policies. Similarly, in the Chinese survey, as the response-length cutoff for defining compliance with the perspective-taking treatment increases, the effect estimate for respondents exposed to the US de-escalation treatment becomes larger, as is the case with the ratio estimator for respondents exposed to the US escalation treatment. Noncompliance thus makes the ATE results reported in the main text conservative estimates. Importantly, the estimates in Tables 3-4 assume one-sided noncompliance. It is also possible for there to be two-sided noncompliance, in that some participants in the control condition may already be in engaged in perspective-taking regardless of treatment assignment; in a potential outcomes framework, these would be “always takers”. However, in this case the ratio estimator provides a lower bound on the effect, since non-compliance in the control group decreases the size of the compliance group (reflected in the denominator), increasing the estimated CACE, such that the results presented here remain conservative tests.

The potential for some respondents to engage in perspective taking, even when they are in the control condition deserves further attention. In a potential outcomes framework, these would be the “always takers.” Indeed, we do find some evidence of perspective taking in the control condition, but not in the same manner induced by the treatment condition. The STM results in section 3.2.1 of the paper show that Chinese respondents in the US de-escalates condition are less likely to think about the strategic calculations of neighboring countries once they are asked to engage in perspective taking. For our purposes, this highlights that some people will engage in perspective taking on their own, but that it is not always in a dyadic framework. Additionally, the effect of having “always takers” suggests that our findings are a conservative estimate of the effect of perspective taking.
4.2 Balance tests

Figures 2 and 3 show that randomization was successful, in that respondents in both the Chinese and the US experiments are well-balanced across treatment conditions for a wide range of demographic characteristics.

Figure 2: Balance tests: China experiment

4.3 Exploiting natural variation in perspective-taking motivation

Although IR scholars tend to view perspective taking as having powerful palliative effects, it is worth noting that many of us aren’t very good at perspective taking. Many of the psychological biases depicted by models of bounded rationality show how poorly we tend to engage in perspective taking. Attribution biases, for example, are biases in person perception (Ross, 1977), ingroup disconfirmation biases mean we reject schema-inconsistent information about others (Larson, 1994), ethnocentric biases mean we differentially treat members of ingroups and outgroups (Kinder and Kam, 2010), and anchoring and adjustment biases mean that when most of us try to imagine how others feel in a given situation, we do so by imagining how we would feel in that situation, and then adjust outwards – usually insufficiently (Epley et al., 2004).

The results from the paper only look at the effect of perspective-taking when it is experimentally induced, but we also know that perspective-taking is an individual difference; it is a faculty that some respondents are naturally better at employing than others (Davis, 1983; Baron-Cohen and Wheelwright, 2004; Wakabayashi et al., 2006; Konrath, O’Brien and Hsing, 2011). We examine how this natural variation in propensity to engage in perspective-taking is associated with escalation preferences, while also acknowledging that these results should be approached with caution given recent questions raised by Murphy and Lilienfeld (2019) about the validity of the self-reported measures of

---

2While some political scientists argue that leaders are more strategically skilled — and thus, presumably, better at perspective taking — than ordinary citizens, the psychological literature finds that individuals in positions of power are actually worse at perspective-taking (Galinsky et al., 2006)
cognitive perspective-taking. Does perspective-taking have palliative effects when studied dispositionally, rather than induced experimentally? The top half of Table 5 presents results from a series of regression models in the US sample, and the bottom half for the China sample, both of which estimate the effect of dispositional levels of perspective-taking controlling for the perspective-taking treatment and a variety of dispositional variables.

For the US sample when China de-escalates, dispositional perspective-taking has a negative sign as shown in models 1, 2, and 3 of Table 5. Although none reaches significance, the negative sign is consistent with the experimental effects, in which US respondents in the PT treatment are more likely to reciprocate de-escalation with de-escalation. Similarly, we find that the one significant effect of dispositional perspective-taking in the US sample is when China escalates, which once again mirrors the experimental results, whereby the US respondents in the PT treatment were more likely to reciprocate escalation with escalation. Thus, although the significance of the results varies, the general direction of the the effects of experimentally induced and dispositional perspective-taking are consistent.

The dispositional results also shed light on one of our experimental findings. In the paper, we found that the perspective-taking treatment did not have a significant effect in the US sample in the experimental condition where the other side escalates. The significant negative interaction term in model 6 of Table 5 suggests an interesting explanation for this effect. Substantively, the perspective-taking treatment has no significant effect on policy preferences among those respondents who are naturally predisposed to engage in perspective-taking, but a strong positive effect among respondents who are dispositionally low in perspective-taking. When the other side escalates, only those Americans least prone to engage in perspective-taking naturally need to be encouraged to do so. This is consistent with Galinsky et al. (2006), who find that power and perspective-taking are negatively correlated: when the US is in a position of relative strength in the region, respondents are less likely to naturally engage in perspective-taking unless encouraged to do so, whereas when the tables turn, only the least empathetic respondents need to be encouraged to engage in perspective-taking. Moreover, the sign of the effect is consistent with the reciprocity pattern reported in the US results above, since for those low-perspective-taking respondents for whom the treatment has a significant effect, the effect is positive, mirroring Chinese behavior.

Next we consider the results for Chinese respondents and find that dispositional perspective-taking is positively signed when the US de-escalates, as is shown in models 1, 2, and 3 of Table 5. The positive sign, although not significant, is consistent with the experimental findings, where Chinese
Table 5: The dispositional perspective-taking results offer little support for the palliative effects of perspective-taking hypothesis

<table>
<thead>
<tr>
<th></th>
<th>US de-escalates</th>
<th>China de-escalates</th>
<th>China escalates</th>
<th>China escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Dispositional empathy</td>
<td>−0.201</td>
<td>−0.087</td>
<td>−0.092</td>
<td>−0.097</td>
</tr>
<tr>
<td></td>
<td>(0.298)</td>
<td>(0.278)</td>
<td>(0.382)</td>
<td>(0.313)</td>
</tr>
<tr>
<td>Perspective taking</td>
<td>−0.268**</td>
<td>−0.204*</td>
<td>−0.211</td>
<td>−0.056</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.107)</td>
<td>(0.425)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>2.473***</td>
<td>2.473***</td>
<td>3.352***</td>
<td>3.318***</td>
</tr>
<tr>
<td></td>
<td>(0.309)</td>
<td>(0.309)</td>
<td>(0.297)</td>
<td>(0.296)</td>
</tr>
<tr>
<td>National chauvinism</td>
<td>0.786***</td>
<td>0.786***</td>
<td>0.915***</td>
<td>0.940***</td>
</tr>
<tr>
<td></td>
<td>(0.213)</td>
<td>(0.213)</td>
<td>(0.215)</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.273**</td>
<td>0.273**</td>
<td>0.014</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.109)</td>
<td>(0.109)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.073</td>
<td>0.073</td>
<td>−0.166</td>
<td>−0.163</td>
</tr>
<tr>
<td></td>
<td>(0.220)</td>
<td>(0.221)</td>
<td>(0.217)</td>
<td>(0.216)</td>
</tr>
<tr>
<td>Dispositional empathy x PT</td>
<td>0.009</td>
<td></td>
<td></td>
<td>−1.561***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.553)</td>
<td></td>
<td>(0.552)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.518***</td>
<td>2.432***</td>
<td>4.835***</td>
<td>2.375***</td>
</tr>
<tr>
<td></td>
<td>(0.234)</td>
<td>(0.296)</td>
<td>(0.249)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>N</td>
<td>720</td>
<td>703</td>
<td>703</td>
<td>773</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.005</td>
<td>0.165</td>
<td>0.164</td>
<td>−0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>US de-escalates</th>
<th>US escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Dispositional Empathy</td>
<td>0.422</td>
<td>−0.185</td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
<td>(0.311)</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>0.248**</td>
<td>0.216**</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>2.129***</td>
<td>2.109***</td>
</tr>
<tr>
<td></td>
<td>(0.307)</td>
<td>(0.308)</td>
</tr>
<tr>
<td>National chauvinism</td>
<td>0.403*</td>
<td>0.409**</td>
</tr>
<tr>
<td></td>
<td>(0.206)</td>
<td>(0.206)</td>
</tr>
<tr>
<td>Education</td>
<td>0.447</td>
<td>0.446</td>
</tr>
<tr>
<td></td>
<td>(0.360)</td>
<td>(0.360)</td>
</tr>
<tr>
<td>Male</td>
<td>0.081</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>Party Member</td>
<td>0.350***</td>
<td>0.349***</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Dispositional empathy x PT</td>
<td>−0.672</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.607)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.865***</td>
<td>3.111***</td>
</tr>
<tr>
<td></td>
<td>(0.237)</td>
<td>(0.354)</td>
</tr>
<tr>
<td>N</td>
<td>754</td>
<td>746</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.006</td>
<td>0.101</td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01
respondents in the perspective-taking treatment were more likely to respond to de-escalation with escalation. This means that in three of the four experimental settings the general effects of dispositional perspective-taking are in the same direction as the effects of the experimental perspective-taking treatment. However, in the fourth combination, which is the Chinese sample when the US escalates, the effect of dispositional perspective-taking diverges from the effect of the experimental perspective-taking treatment. In this case, dispositional perspective-taking is positively signed, whereas the PT treatment had a negative effect. While this one difference may be concerning, it is not robust to controlling for threat perception or once the interaction effect of the perspective-taking treatment and dispositional perspective-taking is included.

4.4 Structural topic model results in the escalation conditions

The results in the main text present the STM results for China and the United States when the other side de-escalates, which is the most theoretically interesting pairing because it sheds light on the mechanisms driving the diverging effects of perspective taking between the two samples. However, we also estimated STM results for the Chinese and American samples when the other side escalates. In the US sample, when China escalates, there are no significant effects of the perspective taking treatment on topical prevalence, consistent with the perspective taking treatment not displaying a significant treatment effect on policy preferences in the US sample when China escalates.

In the escalation condition in Figure 4, we find that the perspective taking treatment significantly affects Chinese participants’ policy explanations, making respondents more likely to to mention Chinese intrinsic interests, repeatedly mentioning the extent to which the conflict is taking place at China’s “doorstep”, or invoking its maritime interests in the region. This suggests that one reason why Chinese respondents de-escalate in the perspective taking condition when faced with US escalation is because perspective taking causes them to perceive Chinese interests as sufficiently and self-evidently high that such costly actions are unnecessary.
4.5 Rumination Effects

To test whether a rumination effect is a leading mechanism underpinning the effect of perspective taking, we proxy for rumination using response latency (Mulligan et al., 2003): that is, how long respondents spent when engaging with the scenario, under the assumption that respondents with higher response latencies engaged in more rumination than respondents with lower response latencies. We therefore calculate respondents’ response time for the screen of the study where respondents were shown the scenario, and administered the two dependent variables of interest (their policy preferences, and their open-ended response justifying their policy choice).

If perspective-taking is simply akin to rumination, it should be the case that respondents who did not receive the perspective-taking treatment, but who display a long (e.g. slow) response latency, should express policy preferences more similar to respondents who did receive the perspective-taking treatment. We therefore calculate the natural log of respondents’ response time on the screen, to account for outliers, and regress respondents’ policy preferences on their logged response times, among those respondents who were given the perspective-taking control. We do so for each of the three escalation conditions (China de-escalates in the US sample, US de-escalates in the Chinese sample, US escalates in the Chinese sample) where the perspective taking treatment has a statistically significant effect.

In the US sample, in the China de-escalates condition, the effect of response time ($B = -0.07$ in a bivariate model, $B = -0.01$ in a multivariate model also controlling for basic demographic characteristics) is not statistically significant ($p < 0.35; p < 0.91$).

In the Chinese sample, in the US de-escalates condition, the effect of response time ($B = -0.31$ in a bivariate model, $B = -0.29$ in a multivariate model also controlling for basic demographic characteristics), is statistically significant ($p < 0.01, p < 0.02$), but the sign is in the wrong direction: that is, whereas the perspective taking treatment makes Chinese respondents more likely to want to escalate in the US de-escalates condition, respondents with higher response latencies in the control condition are more likely to want to de-escalate. As before, then, respondents in the perspective-taking control engaged in more rumination are not more likely to resemble respondents in the perspective taking condition.

In the Chinese sample, in the US escalates condition, the effect of response time ($B = -0.15$ in both a bivariate model and a multivariate model also controlling for basic demographic characteristics), is statistically significant ($p < 0.09, p < 0.09$) and in the right direction, in that like the perspective taking treatment in the US escalates condition, respondents with higher response latencies are more likely to want to de-escalate. However, the fact that response latency has a similar effect in the US de-escalates condition as in the US-escalates condition suggests that this pattern is due to a general association between response latencies and hawkish escalatory preferences in the Chinese sample.

Together, then, this analysis suggests that the perspective taking treatment is more than simply a rumination effect.

4.6 Follow-up attribution experiment: Hawaii

Our follow-up study was fielded in the autumn of 2019 on a convenience sample of $N=627$ American adults recruited through Amazon Mechanical Turk. In the follow-up study, we replicated the attribution module from the first study, but this time set the potential dispute in maritime zones off the coast of Hawaii. Since the MTurk respondents differ from national population parameters in a variety of ways (Huff and Tingley, 2015), such that the sample differs in its demographic composition from the 2016 experiment, we use entropy balancing (Hainmueller, 2012) to reweight the 2019 sample to match the demographic and partisan distributions of the 2019 sample.

To test how much of the difference in attribution asymmetries between the Chinese and Amer-

---

3Following Burleigh, Kennedy and Clifford (2018), we drop respondents using Virtual Private Servers (VPS) to mask their location, or whose locations we could not validate as being in the United States, leaving an effective sample size of $N = 627$.

4See Kertzer, Renshon and Yarhi-Milo (2021) for a similar approach.
icans is due to the location of the scenario and associated geo-strategic concerns, we compare the magnitude of the differences in attribution asymmetries from the solid and dashed blue point estimates in Figure 2 in the main text. The results show that when the dispute is taking place close to Hawaii rather than in the South China Sea, Americans perceive Chinese escalation as significantly more offensively-motivated \((p < 0.001)\). We find that the Chinese attribution asymmetry was 3.1 times greater than the American asymmetry in the first study (solid-red compared to solid-blue), but the Chinese asymmetry is only 2.2 times larger when the American attributions are made in their own backyard (solid-red compared to dashed-blue). This analysis shows that a substantial proportion of the attribution asymmetry is driven by the geography of the dispute and its associated meanings to each population; however, the Chinese attribution asymmetry remains significantly larger than the Americans’, suggesting that the stronger Chinese security dilemma thinking detected in the first study is not merely an artifact of geographic proximity.

Strikingly, however, Americans don’t see US escalation as significantly more defensively motivated in one context than the other \((p < 0.20)\): that is to say, Americans perceive US escalation in China’s backyard as just as defensively motivated as when US escalation occurs close to American territory; Americans perceive American escalation to be defensively motivated regardless of where it’s located. So, although the magnitude of the attribution asymmetry Americans report is significantly larger in the Hawaii experiment than the China experiment \((t = -2.89, p < 0.004)\), the difference is due to changes in the attributions for Chinese behavior, rather than changes in self-attribution. That Americans appear to take the defensive nature of American escalation to be self-evident is a further illustration of the psychological microfoundations of security dilemma thinking at work.

References

Huff, Connor and Dustin Tingley. 2015. ““Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents.” *Research & Politics*


