Perspective Taking and the Security Dilemma: Cross-National Experimental Evidence from China and the United States

Joshua D. Kertzer, Ryan Brutger, and Kai Quek

November 26, 2019

Word count: 11469 words

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 defensively-motivated reasons, failing to realize how it is perceived by the other side. Yet how susceptible to security dilemma thinking are the Chinese and American publics? Can its deleterious effects be mitigated? We explore the individual-level microfoundations of security dilemma thinking, fielding a unique parallel dyadic cross-national survey experiment in China and the United States. We find micro-level evidence consistent with the logic of the security dilemma, and show it is especially pronounced among Chinese respondents. We also find that IR scholars have overstated the palliative effects of perspective taking: perspective taking significantly affects respondents’ policy preferences, but in the face of perceived threats to actors’ identities and goals can lead to a desire for escalation rather than cooperation. Our findings have important implications for the study of public opinion in China-US relations, and perspective taking in IR.

1Paul Sack Associate Professor of Political Economy, Department of Government, Harvard University. Email: jkertzer@gov.harvard.edu. Web: http://people.fas.harvard.edu/~jkertzer/

2Assistant Professor, Department of Political Science, University of California Berkeley. Email: brutger@berkeley.edu. Web: https://sites.google.com/berkeley.edu/brutger/.

3Assistant 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, failing 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 microfoundations of perspective taking in the security dilemma. We are interested in two questions. First of all, in an era when pundits routinely express concern about publics in the two countries being on a collision course with one another, how susceptible are the two countries to security dilemma thinking? 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 the security dilemma’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 that constitutes security dilemma thinking: 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. 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. In echoing other recent research questioning long-standing assumptions about the putatively liberal nature of mass public preferences (Sagan and Valentino, 2017), 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 voluminous body of work on 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 over the South and East China Seas. 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 have tended to emphasize perspective taking’s palliative effects, it can also have a dark side, particularly when actors perceive threats to their identities and goals. We then discuss the mechanics of our cross-national experiments, before presenting our findings.

1 The security dilemma in US-China relations

Policy-makers 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 historically tended to be periods of instability (Kugler and Lemke, 2000; Allison, 2017). Importantly,

---

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/itpir/_documents/trip/ivory_tower_survey_2012.pdf.
however, the dire predictions of power transition theory are contingent upon the rising power being a revisionist state, seeking to overturn the established order. 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 to refer to a wide range of phenomena, but for our purposes, it begins 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, Japan and South Korea — thus loom large in much of the East Asian security scholarship (e.g. Christensen, 1999, 2002; Liff and Ikenberry, 2014), 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).²

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 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 recipro-

cal 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. The security dilemma is characterized by an asymmetric attribution pattern we call *security dilemma thinking*, in which actors attribute offensive motivations to others for behavior they would recognize as defensively motivated if carried out by their own side. This asymmetric attribution thus makes actors more likely to want to escalate, leading to a potential spiral of conflict.

IR scholars have pointed to two potential sets of solutions to these attributional challenges. The first, frequently emphasized in rationalist models, involves costly signals of reassurance, in which states carry out actions that only defensively-motivated states would take in order to convince others of their trustworthiness (Kydd, 2005). Yet models of costly signaling assume that the meaning of signals are obvious to observers, and psychological research suggests that signals may not always speak for themselves (Kertzer, Rathbun and Rathbun, 2019). The second, looming large in psychological models of the security dilemma (e.g. Jervis, 1978; Booth and Wheeler, 2008; Baker, Forthcoming), and our focus here, concerns perspective taking.

### 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.³ 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),

---

³Perspective 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 empathy as a specific subtype of empathy (Davis, 1980), and others still see empathy as a specific subtype of perspective taking (Mohr et al., 2013). We sidestep these debates for our purposes, apart from noting the connections between the two constructs.
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 interna-
tional 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 (Fearon, 1995).

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
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 insuf-
ficient 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; Simonovits, Kézdi
and Kardos, 2018; 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 under-
standing; 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
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 dilemma thinking 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 (Forthcoming, 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 misperceive others’ actions as 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 accurate (Epley and Caruso, 2008). Most relevant for our purposes, however, is that perspective taking also has a dark side (Sassenrath, Hodges and Pfattheicher, 2016). Cognitively, perspective taking causes us to activate knowledge structures both about ourself 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 three ingredients often present in IR.4

The first is negative metastereotypes. When actors engage in perspective taking in intergroup contexts, the first thing they “are apt to see when they try to look through an out-group member’s eyes is themselves” (Vorauer and Sasaki, 2009, 191). Perspective-takers thus activate metastereotypes, beliefs about how they imagine the outgroup views them. If those metastereotypes are

---

4 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).
negative — as would be the case in enduring rivalries, or for actors with historical memory of being targeted by the outgroup — perspective taking’s effects will likely be negative as well.

The second 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 third 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). In an innovative set of experiments, Caruso, Epley and Bazerman (2006) (see also Epley, Caruso and Bazerman, 2006) studied perspective taking in 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, 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; Stein, 2013), 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.

Together, these considerations suggest a number of important implications for the effects of perspective taking in averting potential security dilemmas between the United States and China. Most importantly, because the effects of perspective taking depend on the knowledge structures it activates, perspective taking will have differing effects for citizens of each country. In particular, when actors are engaging in perspective taking in regards to disputes in the South China Sea, the objects under dispute have very different meanings to participants on either side. Since Americans do not view the South China Sea as part of the United States, and there is no interest in taking the territory, Americans’ strategic goal in the region is largely one of maintaining the status quo, and preventing conflict from breaking out. Americans thus perceive little threat to their identity in the dispute and few threats to their goals. For Americans, then, engaging in perspective taking will emphasize positive goal interdependence, which should promote reciprocity and cooperative policy preferences. In contrast, the Chinese view the South China Sea as a rightful part of China’s territory that needs to be protected, particularly if it can be done so while avoiding direct conflict. Chinese thus perceive greater threats to their identity in the dispute, as well as threats to their goals. For Chinese respondents, then, engaging in perspective taking will emphasize negative goal interdependence, which induces them to assert their territorial claim and pursue escalatory policies, but may also activate a heightened sense of caution given their traditional self-perception as being
relatively weaker than the United States (Zhang, 2005).

Importantly, our theory of perspective taking’s contingent effects acknowledges the role of both cognition and culture. 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; Chopik, O’Brien and Konrath, 2017), 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 look for evidence of security dilemma thinking, and explore the potential effects of perspective taking in the context of US-China relations, 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 the psychological microfoundations of security dilemma thinking in the mass public, rather than elite decision-makers. 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 security dilemma thinking 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 (Weiss, 2013, see also Gries, 2001; Wallace and Weiss, 2015). 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 monitoring public opinion; 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 Quek, 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 1,822 American adults recruited by Survey Sampling International (SSI), 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). 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) matching 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 the 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

5 For other recent experiments in political science fielded on SSI, see e.g. Brutger and Kertzer (2018); Quek (2017). For other recent survey experiments in China, see Huang (2015); Bell and Quek (2018); Quek and Johnston (2017/18); Weiss and Dafoe (2019).
6 For more on how our study addressed social desirability bias, see our discussion in Appendix §1.
7 For more details on the survey instrument and translation procedures, see Appendix §2.
Figure 1: Study design

**Experimental modules**

- **Introduction**
  - Describe tensions in South China Sea

- **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?
  - Manipulate perspective-taking
    - Control
    - Writing task I: What conclusion would your side draw?

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

**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)

**Order manipulation**
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
module, and the perspective taking experiment. The objective and setup of each module are somewhat
different. The attribution module 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")). The purpose of this module is to test for the asymmetrical attribution pattern we
associate with security dilemma thinking: 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 (Fearon, 1994), 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 (e.g. Tomz, 2007; Trager and Vavreck, 2011; Kertzer
and Brutger, 2016). 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. 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.
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 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.\textsuperscript{8} Whereas the purpose of the attribution module is to test for security dilemma thinking, the objective of the perspective taking experiment is to see whether perspective taking can mitigate the escalatory dynamics that follow it.

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 involving a hypothetical territorial dispute taking place off the shores of Hawaii, which we describe in greater detail in Appendix \S5.

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

\textsuperscript{8}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 \S4.1, showing it renders the results reported in the main text a more conservative test.
1. By comparing the attributions respondents offer for Chinese and American escalation in the attribution module, we can test the microfoundational assumption of the security dilemma by measuring propensity for *security dilemma thinking*, 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.

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.

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.

4. 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.

### 3 Results

We structure our presentation of the results around the four quantities of interest enumerated above. We begin with the attribution module, looking for micro-level evidence consistent with security dilemma thinking 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, especially in China. Having shown that security dilemma thinking exists, we then turn to the perspective taking experiment, showing how perspective taking encourages reciprocity among American respondents, but opportunistic behavior among Chinese respondents; 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 Asymmetric attributions and 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.

Figure 2: Respondents display the attribution asymmetry consistent with security dilemma thinking

![Figure 2](image)

**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. The figure pools across treatments; in the supplementary analysis we also condition on both the perspective taking treatment and the original escalation treatment, although the results remain the same.

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 and Chinese. 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
nation is acting defensively, but both also believe the other side is acting offensively, likely fostering greater fears among domestic audiences and increased support for tit-for-tat escalation. Interestingly, however, the magnitude of the attribution asymmetry is about three 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. In Appendix §5 we use a follow-up experiment to show that these cross-national differences in the magnitude of the attribution asymmetry are partially attributable to the differential proximity of the dispute to each side; relocating the dispute to maritime zones off the shores of Hawaii makes US respondents more closely — although not completely — resemble the Chinese respondents in the first study. Strikingly, Americans don’t see US escalation as significantly more defensively motivated in one context than the other, offering further evidence of security dilemma thinking. 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. 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 relationship between perspective taking and the security dilemma, 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 3 presents the average policy preferences for respondents in each country, conditioning on treatment assignment.

As Figure 3 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
Figure 3: 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.
counterparts, which is consistent with the conflict’s proximity and perceived threat for the Chinese, a point we return to below.

We only briefly mention the effects of our escalation treatment, since it is of less theoretical interest. By comparing the left and right panels of Figure 3, we find that the public in both countries responds in a manner consistent with the spiral model of escalation 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 $p < 0.001$). This pattern is not surprising, but shows that respondents are reacting in a coherent manner, consistent with pre-existing theories of tit-for-tat escalation.

More important is 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 ($0.23, p < 0.037$). This finding illustrates that perspective taking is not a panacea for alleviating the security dilemma, and 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 ($-0.23, p < 0.017$), suggesting that it makes the Chinese public more likely to be deterred from military escalation when the other side escalates, somewhat mitigating the preference for mutual tit-for-tat escalation.

As noted above, the psychological literature suggests the negative effects we witness in the China experiment when the US de-escalates are especially likely in three conditions: when actors face threats to their identities, when they have negative metastereotypes about the outgroup, and when they perceive threats to their goals or interests. While our research design is not well suited to measuring metastereotypes, in supplementary analyses presented in detail in Appendix §6, we show

---

9Given 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.
that potential disputes in the South China Sea pose lower identity threats to Americans than Chinese respondents, and lower threats to their country’s goals or interests. For example, Chinese respondents are significantly more nationalist than our American respondents are ($t = -3.83, p < 0.001$), and because nationalists are more likely feel threats to their identity when taking the perspective of the outgroup, nationalism also moderates the effect of perspective taking in the Chinese sample: less nationalist respondents are less likely to respond to perspective taking by advocating for escalation when the United States de-escalates, and more nationalist respondents are less likely to respond to perspective taking by advocating for backing down when the United States escalates. Similarly, Chinese respondents perceive greater threats to their goals from the United States than Americans perceive from China ($W = 827560, p < 0.000$), and these differing perceptions of the US-China relationship also interact with the perspective taking treatment: American respondents who perceive China to have negative intentions are less likely to respond to the treatment by reciprocating Chinese de-escalation, while Chinese respondents who perceive the US as having negative intentions are less likely to respond to the treatment by wanting to back down in response to US escalation.

### 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, whose identities and goals are threatened in the dispute, causing them to seek to project strength when the US de-escalates, but de-escalate in the face of American strength.

A similar interpretation can be gleaned from the explanations respondents gave for their preferred policy choice. To do this, 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), a semiautomated text analysis model that allows 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.

While the technical details of STMs are beyond the scope of this paper (see Roberts et al. (2014) for an introduction), they can be thought of as an automated text analysis technique that models
text as a mixture of interpretable “topics.” 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.

Figure 4: Structural topic model results: United States (China de-escalates condition)

![Reciprocity](image)

Results from structural topic models, US sample. The left-hand 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 right-hand panel displays representative responses.

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. Because of space constraints, we focus our results here for respondents in the de-escalation treatment, presenting the results of the escalation treatment in Appendix §4.4.¹⁰

For the American sample in Figure 4, we find that when China de-escalates, the perspective

¹⁰For 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.
taking treatment results in a higher 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, with one respondent writing that the US should “respond positively to China’s willingness to compromise” and another noting that since China decreased its presence the “US could show China’s equal gesture of decreasing naval presence”. Consistent with the above analysis, these findings suggest that the perspective taking treatment makes it more likely that American respondents interpret China’s decrease in military activities as signalling willingness to compromise, which leads them to explain their policy preferences in terms of reciprocal efforts to de-escalate the situation.

Figure 5 displays the STM results for our Chinese sample. It shows that the perspective taking treatment significantly affects Chinese participants’ policy explanations in two different ways. First, consistent with Chinese respondents’ perception of negative goal interdependence and threats to Chinese interests, it increases their emphasis on projecting strength. “The stronger China’s military presence in the South China Sea,” one respondent writes, “the more military force will be deployed by the U.S. in the region, until China’s military power has become so strong that the increase of U.S. strength is no longer meaningful.” Thus, another respondent writes, “China should be neither too weak nor too tough. They will think we are easily bullied if our gesture is too weak, and they will think we want to bully other smaller states if we present too strong a gesture. It is none of the U.S.’s business, they are just meddling around.” The moderate escalation Chinese respondents display when responding to a US de-escalation in the perspective taking condition reflects this logic.

Second, and especially interestingly, asking participants to assess the conclusions the United States would draw leads them to be less likely to think about the strategic calculations of others in the region: participants in the control condition were significantly more likely to discuss the ramifications of China’s behavior on its neighboring countries (offering responses like “Increasing actions will lead to the distrust and suspicion among the neighboring countries, while reducing actions will undermine our control over this area. Thus, the better approach is to maintain the status quo.”). This is helpful in reminding us that although we think about perspective taking in a dyadic manner, in a regional context with multiple actors, investing cognitive resources in thinking about one actor’s perspective can crowd out the perspective of others.

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
Figure 5: Structural topic model results: China (United States de-escalates condition)

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.
on perceived threats to actors’ identities and goals. It makes American respondents, who perceive relatively little threat to their identities or 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 projecting strength (and think less about the effect it will have on neighboring countries) when the US shows signs of weakness. As one Chinese respondent wrote, “The fact that the United States decided to withdraw from the South China Sea issue shows that our country has really become powerful. The United States is no longer the hegemonic power that can intervene with any country and any affair in the world, by whatever measures they want to take! I am really proud to be born as a Chinese!”

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 and misperceptions can be mitigated. In this paper, we explore the role of perspective taking in this process. Given continued tensions in the South China Sea, and concerns about popular nationalism in both the American and Chinese publics, understanding how perceptions on both sides are formed — and shaped by perspective taking — is a useful endeavor for the study of US-China relations.

Fielding a set of cross-national experiments that examine the security relationship between the US and China in a dyadic manner, we find micro-level evidence consistent with security dilemma thinking in both countries’ publics, and that these attribution asymmetries are especially pronounced in China. Despite the volume of literature on China’s rise, and the extent to which IR scholars frequently build their arguments on assumptions about the nationalist content of Chinese public opinion, the longstanding difficulty of obtaining high-quality public opinion data in China has meant that surprisingly little of this literature actually measures this directly.11 Our findings thus 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. Our results show that security dilemma thinking is not merely the preserve of elite decision-makers, but is also detectable in the public writ-large. Indeed, the presence of a substantively large attribution asymmetry and the free responses

11 For a similar critique, see Johnston (2017).
respondents provided show that respondents display many of the key perceptual mechanisms Jervis (2001, 38) associates with the security dilemma: each side is “‘deeply fearful that the other side was aggressive or would become so in the future”, and sees its “own behavior as designed to maintain the status quo”, while the other side attempts to destabilize it. Importantly, for the Chinese, the South China Sea is already a rightful part of China, and others are trying to take it away; for the Americans, the Chinese are expansionist, and each side displays “the fear... of being exploited by the other side.” These findings highlight the potential risks between the US and China in the South China Sea, and the divergent manners in which citizens in each country perceive military signals in the region.

Our study also suggests some skepticism about the palliative effects of perspective taking hypothesis often adopted by IR scholars, showing how perspective taking and threat perceptions shape security dilemma thinking. 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 a 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 highlight conflicts of interest as well. We find that experimentally inducing perspective taking induces people to draw on different knowledge structures, particularly perceived threats to interests and identity, which shape the escalatory preferences of the public and the risks of security dilemma thinking. Perspective taking makes Americans – who perceive fewer threats to their identity and interests – more likely to recognize conciliatory signals and increase their desire to 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.

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, 2017). 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. In this sense, the study differs from
the content-specific perspective taking interventions sometimes used in 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.

References

Adida, Claire L, Adeline Lo and Melina R Platas. 2018. “Perspective taking can promote short-term
inclusionary behavior toward Syrian refugees.” Proceedings of the National Academy of Sciences
115(38):9521–9526.
and Institutions.” World Politics 38(1):58–79.
litical Psychology.
Barth, Markus and Stefan Stürmer. 2016. “Comparison of the Effects of Individual and Group-Level
World Politics. Basingstoke, UK: Palgrave Macmillan.
Brutger, Ryan and Joshua D. Kertzer. 2018. “A Dispositional Theory of Reputation Costs.” Inter-
Butterfield, Herbert. 1951. “The Tragic Element in Modern International Conflict.” The Review of
Caruso, Eugene M., Nicholas Epley and Max H. Bazerman. 2006. “Reactive egoism and unwarranted
distrust.” Unpublished manuscript, Harvard University.
Chen, Dingding, Xiaoyu Pu and Alastair Iain Johnston. 2014. “Debating China’s Assertiveness.”


Holmes, Marcus. 2018. *Face-to-Face Diplomacy: Social Neuroscience and International Relations.*


URL: http://nationalinterest.org/commentary/can-china-rise-peacefully-10204


Yarhi-Milo, Keren. 2014. Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations. Princeton, NJ: Princeton University Press.

Perspective-Taking and the Security Dilemma: Supplementary Appendix
Public version

Contents

1 Note on survey partner .......................... 3
   Table 1: Sample characteristics: Chinese sample ......................... 3
   Table 2: Sample characteristics: US sample ............................... 4

2 Translation protocol for survey instrument .................. 5

3 Survey instrument ........................................ 5

4 Supplementary analysis ................................... 10
   4.1 Causal average complier effects ........................................ 10
      Table 3: Causal average complier effect (CACE) estimates, US sample ... 10
      Table 4: Causal average complier effect (CACE) estimates, Chinese sample . 10
   4.2 Balance tests .................................................. 12
      Figure 1: Balance tests: China experiment .............................. 13
      Figure 2: Balance tests: US experiment .................................. 14
   4.3 Exploiting natural variation in perspective-taking motivation ........ 15
      Table 5: The dispositional perspective-taking results offer little support for the palliative effects of perspective-taking hypothesis ......................... 16
   4.4 Structural topic model results in the escalation conditions .......... 18
      Figure 3: Structural topic model results: United States (China escalates condition) . 19
      Figure 4: Structural topic model results: China (United States escalates condition) . 20

5 Follow-up attribution experiment: Hawaii .................... 21
   Figure 5: Proximity increases the attribution asymmetry .................. 21

6 Explaining cross-national differences ..................... 22
   6.1 Perceived threats to identities ...................................... 23
      Figure 6: Chinese sample significantly more nationalist than US sample ... 24
      Figure 7: Nationalism moderates the impact of perspective taking .................. 25
   6.2 Perceived threats to goals ......................................... 26
      Figure 8: Chinese sample perceives the United States as significantly more threatening 28
      Figure 9: Threat perception interacts with the perspective-taking treatment .... 29
   6.3 Alternative hypotheses ........................................... 30
      Figure 10: Chinese sample not substantially different in k-level reasoning than American sample . 32
      Figure 11: Chinese sample significantly more likely to see military strength as necessary for peace . 33
      Figure 12: Hawkishness does not significantly moderate the impact of perspective-taking 33
      Figure 13: American sample slightly higher in dispositional perspective taking .... 35
      Figure 14: Variation in dispositional perspective taking unlikely to explain cross-national differences .................. 35
6.4 Explaining individual-level variation in attribution asymmetry ........................ 36
Table 6: Attribution Asymmetry, Threat Perception, and Nationalism .................. 38
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.

Table 1: Sample characteristics: Chinese sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>2010 National Census</th>
<th>2016 sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td>Full Population</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>51.2%</td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Northern</td>
<td>12.7%</td>
</tr>
<tr>
<td></td>
<td>Northeastern</td>
<td>8.8%</td>
</tr>
<tr>
<td></td>
<td>Eastern</td>
<td>30.2%</td>
</tr>
<tr>
<td></td>
<td>Central &amp; Southern</td>
<td>27.0%</td>
</tr>
<tr>
<td></td>
<td>Northwestern</td>
<td>14.2%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>≤ 19</td>
<td>24.1%</td>
</tr>
<tr>
<td></td>
<td>20-29</td>
<td>17.1%</td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td>16.1%</td>
</tr>
<tr>
<td></td>
<td>40-49</td>
<td>17.3%</td>
</tr>
<tr>
<td></td>
<td>≥ 50</td>
<td>25.3%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Han</td>
<td>91.6%</td>
</tr>
</tbody>
</table>
Table 2: Sample characteristics: US sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample</th>
<th>Population Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.45</td>
<td>0.49</td>
</tr>
<tr>
<td>Female</td>
<td>0.55</td>
<td>0.51</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>25-44</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>45-64</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>≥ 65</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>0.34</td>
<td>0.42</td>
</tr>
<tr>
<td>Some college/university</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>College/university</td>
<td>0.37</td>
<td>0.28</td>
</tr>
<tr>
<td>Graduate/professional school</td>
<td>0.17</td>
<td>0.10</td>
</tr>
</tbody>
</table>
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

下列问题涉及中国与世界上其他国家的关系。

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

最近南海局势备受关注。区域内的许多国家宣称，他们在有争议的国际水域拥有权利，而这些水域
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

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_

中国国内在讨论，是否应当改变中国在南海的军事存在。
在这种情况下，您认为中国在南海的军事行动应当增加、减少，还是保持不变？
[增加/减少/保持不变]

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

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]
您是否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]
Please briefly explain your response.

ATTENTION MODULE

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.
4 Supplementary analysis

4.1 Causal average complier effects

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

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>China de-escalates</th>
<th>China escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CACE 95% CIs</td>
<td>2SLS estimate</td>
</tr>
<tr>
<td>0</td>
<td>-0.27 (-0.51, -0.05)</td>
<td>-0.27 (-0.5, -0.05)</td>
</tr>
<tr>
<td>5</td>
<td>-0.35 (-0.64, -0.06)</td>
<td>-0.28 (-0.52, -0.05)</td>
</tr>
<tr>
<td>10</td>
<td>-0.36 (-0.67, -0.06)</td>
<td>-0.25 (-0.5, -0.01)</td>
</tr>
<tr>
<td>15</td>
<td>-0.39 (-0.72, -0.07)</td>
<td>-0.26 (-0.53, 0)</td>
</tr>
<tr>
<td>20</td>
<td>-0.42 (-0.77, -0.07)</td>
<td>-0.28 (-0.56, -0.01)</td>
</tr>
<tr>
<td>25</td>
<td>-0.43 (-0.8, -0.08)</td>
<td>-0.33 (-0.62, -0.05)</td>
</tr>
<tr>
<td>30</td>
<td>-0.47 (-0.86, -0.08)</td>
<td>-0.38 (-0.69, -0.08)</td>
</tr>
<tr>
<td>35</td>
<td>-0.51 (-0.95, -0.09)</td>
<td>-0.43 (-0.75, -0.1)</td>
</tr>
<tr>
<td>40</td>
<td>-0.57 (-1.06, -0.1)</td>
<td>-0.41 (-0.76, -0.06)</td>
</tr>
<tr>
<td>45</td>
<td>-0.61 (-1.13, -0.11)</td>
<td>-0.45 (-0.82, -0.08)</td>
</tr>
<tr>
<td>50</td>
<td>-0.66 (-1.25, -0.12)</td>
<td>-0.43 (-0.82, -0.04)</td>
</tr>
</tbody>
</table>

Table 3 presents CACE estimates, defining compliance at different percentile thresholds ($\tau$) of response length. The first two columns in each panel uses the ratio estimator $CACE = \frac{ITT}{P_i(Complied)}$ 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>$\tau$</th>
<th>US de-escalates</th>
<th>US escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ratio estimator</td>
<td>2SLS estimate</td>
</tr>
<tr>
<td>0</td>
<td>0.23 (0.002, 0.44)</td>
<td>0.23 (0.01, 0.45)</td>
</tr>
<tr>
<td>5</td>
<td>0.24 (0.002, 0.45)</td>
<td>0.25 (0.03, 0.47)</td>
</tr>
<tr>
<td>10</td>
<td>0.24 (0.002, 0.45)</td>
<td>0.25 (0.03, 0.47)</td>
</tr>
<tr>
<td>15</td>
<td>0.25 (0.002, 0.48)</td>
<td>0.38 (0.14, 0.62)</td>
</tr>
<tr>
<td>20</td>
<td>0.25 (0.002, 0.48)</td>
<td>0.38 (0.14, 0.62)</td>
</tr>
<tr>
<td>25</td>
<td>0.28 (0.003, 0.54)</td>
<td>0.46 (0.19, 0.72)</td>
</tr>
<tr>
<td>30</td>
<td>0.31 (0.003, 0.59)</td>
<td>0.52 (0.24, 0.81)</td>
</tr>
<tr>
<td>35</td>
<td>0.32 (0.003, 0.61)</td>
<td>0.56 (0.27, 0.85)</td>
</tr>
<tr>
<td>40</td>
<td>0.36 (0.003, 0.68)</td>
<td>0.47 (0.16, 0.79)</td>
</tr>
<tr>
<td>45</td>
<td>0.39 (0.004, 0.74)</td>
<td>0.41 (0.07, 0.74)</td>
</tr>
<tr>
<td>50</td>
<td>0.43 (0.004, 0.82)</td>
<td>0.35 (0, 0.71)</td>
</tr>
</tbody>
</table>

Table 3 presents CACE estimates, defining compliance at different percentile thresholds ($\tau$) of response length. The first two columns in each panel uses the ratio estimator $CACE = \frac{ITT}{P_i(Complied)}$ 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
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 1 and 2 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 1: Balance tests: China experiment
Figure 2: Balance tests: US experiment

- Age
- Gender
- Party ID
- Household income
- Race
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.\(^1\) 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 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

\(^1\)While 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)
Table 5: The dispositional perspective-taking results offer little support for the palliative effects of perspective-taking hypothesis

<table>
<thead>
<tr>
<th>Sample: US</th>
<th>China de-escalates</th>
<th>China escalates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Dispositional perspective-taking</td>
<td>$-0.201$</td>
<td>$-0.031$</td>
</tr>
<tr>
<td></td>
<td>$(0.298)$</td>
<td>$(0.278)$</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>$-0.268^{**}$</td>
<td>$-0.210^{*}$</td>
</tr>
<tr>
<td></td>
<td>$(0.116)$</td>
<td>$(0.108)$</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>$2.640^{***}$</td>
<td>$2.640^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.301)$</td>
<td>$(0.301)$</td>
</tr>
<tr>
<td>National chauvinism</td>
<td>$0.644^{***}$</td>
<td>$0.644^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.234)$</td>
<td>$(0.235)$</td>
</tr>
<tr>
<td>Education</td>
<td>$0.268^{**}$</td>
<td>$0.210^{*}$</td>
</tr>
<tr>
<td></td>
<td>$(0.116)$</td>
<td>$(0.108)$</td>
</tr>
<tr>
<td>Male</td>
<td>$0.235^{**}$</td>
<td>$0.235^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.110)$</td>
<td>$(0.110)$</td>
</tr>
<tr>
<td>Ideology</td>
<td>$0.129$</td>
<td>$0.129$</td>
</tr>
<tr>
<td></td>
<td>$(0.220)$</td>
<td>$(0.220)$</td>
</tr>
<tr>
<td>Dispositional PT x PT</td>
<td>$-0.008$</td>
<td>$-1.403^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.554)$</td>
<td>$(0.552)$</td>
</tr>
<tr>
<td>Constant</td>
<td>$4.518^{***}$</td>
<td>$2.316^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.234)$</td>
<td>$(0.306)$</td>
</tr>
<tr>
<td>N</td>
<td>720</td>
<td>704</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.005</td>
<td>0.159</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample: China</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 perspective-taking</td>
<td>$0.422$</td>
<td>$0.016$</td>
</tr>
<tr>
<td></td>
<td>$(0.317)$</td>
<td>$(0.315)$</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>$0.248^{**}$</td>
<td>$0.214^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.111)$</td>
<td>$(0.109)$</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>$1.123^{***}$</td>
<td>$1.112^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.229)$</td>
<td>$(0.230)$</td>
</tr>
<tr>
<td>National chauvinism</td>
<td>$0.062^{**}$</td>
<td>$0.061^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.030)$</td>
<td>$(0.030)$</td>
</tr>
<tr>
<td>Education</td>
<td>$0.617^{*}$</td>
<td>$0.612^{*}$</td>
</tr>
<tr>
<td></td>
<td>$(0.369)$</td>
<td>$(0.369)$</td>
</tr>
<tr>
<td>Male</td>
<td>$0.116$</td>
<td>$0.117$</td>
</tr>
<tr>
<td></td>
<td>$(0.110)$</td>
<td>$(0.110)$</td>
</tr>
<tr>
<td>Party Member</td>
<td>$0.350^{***}$</td>
<td>$0.349^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.122)$</td>
<td>$(0.122)$</td>
</tr>
<tr>
<td>Dispositional PT x PT</td>
<td>$-0.781$</td>
<td>$-0.119$</td>
</tr>
<tr>
<td></td>
<td>$(0.619)$</td>
<td>$(0.095)$</td>
</tr>
<tr>
<td>Constant</td>
<td>$4.865^{***}$</td>
<td>$3.402^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.237)$</td>
<td>$(0.399)$</td>
</tr>
<tr>
<td>N</td>
<td>754</td>
<td>747</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.006</td>
<td>0.065</td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01
general direction of 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 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

Due to space constraints, 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. Figure 3 presents the results in the American sample when China escalates, showing that even though China is escalating, the perspective taking treatment nonetheless causes respondents to perceive greater positive goal interdependence with China. The perspective taking treatment is associated with a significant decrease in language emphasizing Americans’ perceived threat and concern for protecting other countries. Encouraging our American respondents to take the perspective of the other side appears to make them less likely to perceive Chinese actions as threatening, which reduces the likelihood that people believe we need to protect ourselves or others from Chinese actions. Importantly, the STM recovers a significant decrease in this topic in the PT condition even though the PT treatment does not display a significant treatment effect on policy preferences. This pattern is likely due to what the causal inference literature refers to as an “inconsistent mediation effect”, in which a treatment affects multiple mechanisms of different sign, such that they cancel one another out, producing a non-significant total effect (MacKinnon, Krull and Lockwood, 2000). This would be the case if the perspective taking treatment decreases threat perception in the face of escalation, while also increasing other factors that have countervailing effects on policy preferences.

In the escalation condition in Figure 4, we find that the perspective taking treatment significantly affects Chinese participants’ policy explanations along three different dimensions. First, the perspective taking treatment decreases the extent to which participants emphasize increased military action. When faced with US escalation, respondents in the perspective taking treatment are less hawkish than respondents in the control (and thus less likely to offer statements like “The military has no field experience, and military operations could help to enhance its combat capability”, or “The situation in the South China Sea is getting serious, China is facing too many opponents in this case, thus it needs to rapidly increase its military strength and actions.”) Second, respondents in the perspective taking treatment are less likely to offer nationalist arguments emphasizing the inviolability of Chinese sovereignty, and thus less likely to make claims like “China shall not lose
Results from structural topic models, US sample. The left-hand 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 right-hand panel displays representative responses.

any one inch of the territory, and whoever sells any bit of the territory in exchange for a moment of peace is the sinner of the Chinese nation and history", or “The Chinese have been silent for too long, and we will go to war for the country and national dignity!” Finally, respondents in the perspective taking condition are more likely to mention Chinese intrinsic interests, repeatedly mentioning the extent to which the conflict is taking place at China’s “front door”, 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 signals are unnecessary.
The situation at the South China Sea is getting serious, China is facing too many opponents in this case, thus it needs to rapidly increase its military strength and actions. The military now has no field experience, and military operations could help to enhance its combat capability.

China should increase military actions to show our country's comprehensive national strength and military strength and put pressures on the United States and other countries.

For now, some increase in military operations could protect our national interests.

China always stands for peace and insists on peaceful foreign policies, but certain increase in military operations is conducive to maintaining China's majesty and a peaceful environment.

The Chinese nation is a benevolent nation, but it does not mean that our tolerance is unlimited! We always use force only after showing our courtesy, but it does not imply that we will remain silent even if you are bullying us. Our people have an unyielding spirit, with strong determination. Any country which intervenes the territorial integrity and sovereignty of China will be taken down no matter how far it is from. The Chinese have been silent for too long, and we will go to war for the country and national dignity!

China shall not lose any one inch of the territory, and whoever sells any bit of the territory in exchange for a moment of peace is the sinner of the Chinese nation and history.

Defend the national sovereignty and territorial integrity.

The threats cannot be resisted if there is not enough military force to counter such threat.

No wonder that the U.S. will come all the way over to fight for their so-called "interests" in the region when China is not even able to defend its South China Sea which is its home door.

Natural and fishery resources.

Right at the front door.

China has restrained itself to its utmost in the South China Sea maritime rights disputes. Even though individual southeast Asian countries (the Philippines, Vietnam) have recurrently, illegally occupied China's reefs in the south China sea and natural resources in a large-scale, China still suggests peacefully solving the problems rather than resorting to force.
5 Follow-up attribution experiment: Hawaii

There are two interpretations of the differential attribution asymmetries we report in the main text. The first is that Chinese respondents are especially susceptible to security dilemma thinking; the second is that this asymmetry is due to American (de)escalation meaning something very different in China’s backyard than Chinese (de)escalation does in the same context.

We thus fielded a follow-up study on in the autumn of 2019 on a convenience sample of N=627 American adults recruited through Amazon Mechanical Turk. In this 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. Our interest is in how relocating the dispute to a location closer to home changes the attributions Americans make about the scenario. Figure 5 replicates the attribution results from the main study, while also adding results from the Hawaii experiment. 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.

A comparison between the blue and green point estimates in Figure 5 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. Interestingly, however, Americans don’t see US escalation as significantly more defensively motivated in one context than the other: 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. So, although the magnitude of the attribution asymmetry Americans report is significantly steeper 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. These findings therefore offer further evidence of security dilemma thinking. Moreover, although changes in the political situation in China prohibited us from fielding our Hawaii study on a Chinese sample, by comparing the green and red point estimates, we can compare the American sample from the Hawaii study with the Chinese sample from the South China Sea study: the results show that the Chinese attribution

---

2Following 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 $N = 627$.

3See Kertzer, Renshon and Yarhi-Milo (2018) for a similar approach.
asymmetry remains substantially steeper than the American attribution asymmetry, suggesting the results are likely also shaped by respondents’ perceived relationship to the objects of the dispute, in that American ownership of Hawaii is not challenged by others in the way that Chinese ownership of the South China Sea is.

6 Explaining cross-national differences

As noted in the main text, we find strikingly divergent effects of perspective-taking between the American and Chinese samples. The most striking occurs in the de-escalation condition: when the other side de-escalates, perspective taking makes Americans engage in positive reciprocity, but makes the Chinese sample advocate for greater escalation. The psychological research on the dark side of perspective taking suggests that the negative effects we see in the Chinese experiment when the US de-escalates are especially likely in three circumstances: when actors have negative metastereotypes about the outgroup, when actors perceive threats to their identity, and when actors perceive threats to their goals or interests. Our research design isn’t well suited to teasing out and isolating the effects of metastereotypes (that is, what American respondents think Chinese respondents think
about the United States, and vice versa), but there are a number of tests we conduct to explore how perceived threats to actors’ and interests affect perspective taking. We describe each below, along with a number of tests we conduct to rule out alternative hypotheses.

6.1 Perceived threats to identities

The first circumstance when perspective taking can backfire and produce antagonistic rather than palliative effects is when actors perceive threats to their identities. Many models of perspective taking assume a process of self-other overlap in cognitive representations (Galinsky, Wang and Ku, 2008): actors see themselves in others, and see others in themselves. According to some theories of perspective taking (e.g. Cialdini et al., 1997), this form of projection suggests that perspective taking’s prosocial effects are actually egoistically motivated (we treat the target more nicely when we take their perspective not because we’re more altruistic, but because we see ourselves in it, thereby extending to it the self-serving biases we normally apply to ourselves). Importantly, however, when actors with strong emotional attachments to their ingroup engage in perspective taking with an outgroup, it can constitute a threat to their social identity and lead to outgroup derogation (e.g. Tarrant, Calitri and Weston, 2012; Sassenrath, Hodges and Pfattheicher, 2016; see also Berndsen, Thomas and Pedersen, 2018).

Although there are a variety of mechanisms through which identity threats operate in IR (e.g. Mitzen, 2006; Rousseau, 2006), one of the most important is national attachment, reflecting the extent to which respondents identify with their country (Herrmann, Isernia and Segatti, 2009). 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 to have a starker perception of shared fate (Brewer and Brown, 1998). 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 test the role that within- and between-sample variation in nationalism plays in our results, we administer measures of national attachment to both our American and Chinese samples. We measure national attachment in the United States with two items from the national attachment scale of Herrmann, Isernia and Segatti (2009) (“When someone says something bad about the American people, how strongly do you feel it is as if they said something bad about you?” and “How much do you feel that what happens to America in general will be your fate?”); for our Chinese sample, we use the same question tapping national identification (“When someone says something bad about
the Chinese people, how strongly do you feel it is as if they said something bad about you?"), while also using two standard measures of national pride commonly used in studies of Chinese nationalism ("Are you proud to be a Chinese citizen?", with responses ranging from "Not proud at all" to "Very proud", and "Is China superior to other countries?", with responses ranging from "Not superior" to "Superior by far"; see Johnston (2017) for similar items).

Figure 6 shows that the Chinese sample is significantly more nationalist ($t = -3.83, p < 0.001$) than the American sample; the item-level comparison similarly finds that Chinese respondents are significantly more likely to feel that when someone says something bad about your country, it is as if they said something bad about you, although the Chinese-language survey had a four- rather than five-point set of response options for this item, such that it is also possible that the differences for this item are partially attributable to measurement properties. Importantly, however, we also find in Figure 7 that nationalism also moderates the effect of perspective taking in the Chinese sample: less nationalist respondents are less likely to respond to perspective taking by advocating for escalation when the United States de-escalates, and more nationalist respondents are less likely to respond to perspective taking by advocating for backing down when the United States escalates.

Figure 6: Chinese sample significantly more nationalist than US sample

Note: US sample in blue, Chinese sample in red. The mean of each distribution is represented by a vertical line.

The open-ended responses similarly illustrate the extent to which Chinese respondents perceived

---

4α = 0.5 in the US, α = 0.67 in China

5As is the case with all of the dispositional variables we analyze here, we estimate the marginal effect of the perspective-taking treatment conditional on each trait using a 5-fold least-squares cross-validation kernel estimation procedure introduced by Hainmueller, Mummolo and Xu (2018) in order to flexibly estimate the functional form of the interaction while avoiding imposing a linear functional form.
The figure shows how nationalism moderates the impact of perspective taking in the Chinese sample: less nationalist respondents are less likely to respond to perspective taking by advocating for escalation when the United States de-escalates, and more nationalist respondents are less likely to respond to perspective taking by advocating for backing down when the United States escalates. Each panel displays the marginal effect of the perspective-taking treatment across national attachment, estimated using a kernel estimation procedure with 5-fold least-squares cross-validation using the interflex package in R to avoid making assumptions about the functional form of the interactive effect.
identity threats implicated in the scenario more generally. One respondent wrote: “The Chinese nation is a benevolent nation, but it does not mean that our tolerance is unlimited!... Our people have an unyielding spirit. with strong determination... The Chinese have been silent for too long, and we shall go to war for the country and national dignity.” Another wrote, “China shall not lose any one inch of the territory, and whoever sells any bit of the territory in exchange for a moment of peace is the sinner of the Chinese nation and history.”

6.2 Perceived threats to goals

The second situation where perspective taking can have negative consequences is when actors perceive threats to their goals or motivations. 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). Perceived levels of goal interdependence is an inherent part of the images actors have of one another in international relations; in the most prominent version of image theory in IR, perceived threats or opportunities forms one of the three dimensions of the “combinatorial constructs” actors possess about one another (Boulding, 1959; Herrmann and Fischerkeller, 1995; Herrmann, 2013; Castano, Bonacossa and Gries, 2016). Since threat perceptions in IR are a function of both perceived intentions and perceived capabilities (Stein, 2013), 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.

We measure perceptions of negative goal interdependence using a series of items from image theory. First, we asked our respondents two items getting at perceptions of threat or opportunity: is the other actor a crucial partner for cooperation? Or, a dissatisfied power with expansionist ambitions? To this end, respondents were asked to indicate the extent to which they agree with the statements "[China/the United States] is China’s most important partner", and “[China/the United States] wants to dominate the world", which is our most direct measure of perceived threatening intentions. Second, we administered an item soliciting respondents’ perception of the balance of power between the two countries, asking them the extent to which they agree with the proposition that “With regard to global power, China has closed the gap with the United States". All three Likert responses are scaled such that higher values indicate stronger agreement. These image theory
items are conceptually related to the stereotype content and preexisting perceptions of bilateral relations items that Jing et al. (2017) find exert important influences in their preemptive strike lab experiments among American and Chinese respondents.⁶

In addition to these image theory measures, we also use our attribution measures (described in detail in the main text) as a different way of operationalizing perceived goal interdependence: if you are high in security dilemma thinking, and perceive the other side’s behavior in a given situation as offensively motivated (unlike your own behavior in the same situation), this implies a greater degree of perceived threat than if you see the other actor’s behavior as defensive in origin.

Figure 8 presents differences in perceived negative intentions and relative capabilities between the two countries. The figure shows that Chinese respondents see the United States as significantly more threatening than American respondents see China: perceive greater threats to their goals from the United States than Americans perceive from China. Chinese respondents are significantly more likely to indicate that the United States in harboring threatening intentions than Americans indicate with respect to China ($W = 827560, p < 0.000$), while citizens in the two countries also disagree on their relative power ($W = 1604200, p < 0.000$): Chinese respondents also perceive China as significantly weaker than the United States, while Americans perceive China as more likely to have caught up in power.

As Figure 9 shows, these differing perceptions of the nature of the US-China relationship also interact with the perspective taking treatment: American respondents who perceive China to have negative intentions are less likely to respond to the perspective taking treatment by reciprocating Chinese de-escalation, while Chinese respondents who perceive the US as having negative intentions are less likely to respond to the perspective taking treatment by wanting to back down in response to US escalation. Similarly, the larger the attribution asymmetry — that is, the more perceivers believe the other side to be offensively motivated than one’s own — among the American respondents, the weaker the effect of perspective taking on encouraging reciprocity when the other side backs down; perspective taking displays palliative effects in the US sample only among respondents who perceive China as relatively more defensively motivated than the US. Conversely, when we turn to the Chinese sample in the same experimental condition, we see that the backfire effect of perspective

---

⁶The pre-existing perceptions of bilateral relations that Jing et al. (2017) include, for example, include items measuring the extent to which respondents in each country perceive the bilateral relationship to be competitive or cooperative; similarly, their national stereotype measures of warmth and competence map onto perceived threat, and relative capabilities, as Herrmann (2013) notes.
Figure 8: Chinese sample perceives the United States as significantly more threatening

(a) Perceived intentions and capabilities

(b) Attributions for escalation

Note: US sample in blue, Chinese sample in red. The mean of each distribution is represented by a vertical line.
Figure 9: Threat perception interacts with the perspective taking treatment

(a) Perceived negative intentions

(b) Attribution asymmetry

Note: each panel displays the marginal effect of the perspective-taking treatment across a different relational variable, estimated using a kernel estimation procedure with 5-fold least-squares cross-validation using the interflex package in R to avoid making assumptions about the functional form of the interactive effect.
taking detected among Chinese respondents is only significant among those respondents with large attribution asymmetries.

Importantly, these differences in threat perception also extend beyond the US-China relationship in general, to the setting of the dispute itself. The open-ended responses show that the stakes of the dispute in the South China Sea are low to many Americans. A number of Americans answered the first writing task in a manner that suggested they viewed the dispute through the lens of helping local wildlife, rather than as an urgent security threat, lamenting that China does “not care about the natural habitat”. Regardless of whether China escalated or not, many of the American respondents viewed the scenario through general weariness about the United States being the “world police”, emphasizing that the United States “needs to take care of America’s issues first”, or should simply “STAY OUT OF IT”. In contrast, the Chinese respondents viewed the United States as meddling in core Chinese territory. As one respondent put it, “China and the United States are located on two sides of the planet. On what grounds does the United States have to come all the way over and intervene in our territory?”

6.3 Alternative hypotheses

Although our theory suggests that differential responses to perspective taking between the two samples are likely due to differential threats to respondents’ identities and interests, we also test three alternative hypotheses, each of which points to potential compositional differences between the two samples. The logic behind these hypotheses is that if the two samples differ compositionally from one another in theoretically relevant ways, and those differences interact with the perspective taking treatment, then these compositional differences can help explain why we see cross-national differences in the results.

The first alternative explanation comes from behavioral research on strategic thinking. Behavioral economists often measure strategic thinking through “beauty contest games” (Camerer, Ho and Chong, 2004), based on a discussion in Keynes (1936) comparing the stock market to a beauty contest where players win by accurately guessing which contestants other players will deem to be the most beautiful. In one common version of the game (Coricelli and Nagel, 2009; LeVeck et al., 2014), players are asked to guess a number between 0 to 100, in which the winner will be the player whose guess is the closest to some multiplier (e.g. two-thirds) of the average guess of the group. The guesses are then scored as a measure of k-level reasoning: level-0 reasoners guess randomly between 0-100,
while level-1 reasoners assume the other players will guess randomly such that the best response is 33 \((50 \times \frac{2}{3})\), level-2 reasoners assume other players are capable of \(K - 1\)-level reasoning such that the best response is 22 \((50 \times \frac{2^2}{3})\), and so on. In this sense, k-level reasoning is often used as a measure of “strategic skill” in that level-two reasoners are more strategic than level-one reasoners, and so on (Hafner-Burton et al., 2014). Yet it also measures strategic players’ implicit beliefs about the other players. While the equilibrium of the game is 0, most players in the real world don’t approximate *homo economicus* and iteratively eliminate dominated strategies all the way; winning the game thus requires respondents to engage in strategic perspective taking and attempt to gauge how strategic the other players are. Mindful of these dual meanings, we thus adapted the measure of k-level reasoning used by LeVleck et al. (2014), but with a cross-national twist. For half the respondents in each country, we presented the game as follows:

Pretend you’re playing a game against other players, where every participant has to think of a whole number from 0 to 100. The winner of the game is the player whose guess is 60% of the average guess of the group. (For example, if the average guess of the group is 90, the winner is the participant in the group who guessed the number closest to 54).

What number would you pick?

For the other (random) half of respondents in each country, we repeated the same prompt, but modified the first sentence to begin with “Pretend you’re playing a game against players from China” for the US respondents, or “players from the United States” for the Chinese respondents. By randomizing the identity of the opponents, we can test whether respondents’ k-level reasoning scores differ when playing against opponents from the opposing country. Regardless of treatment condition, we analyze the results the same way, following LeVeck et al. (2014) by differentiating between level-\(K = 0\), \(K = 1\), \(K = 2\), and \(K = 3\) (or greater) respondents.

As it turns out, we find that the identity treatment does not affect respondents’ k-level scores in either country \((p < 0.624\) for the US sample, \(p < 0.858\) for the Chinese sample, each from a Wilcoxon-rank sum test), so in Figure 10 we combine k-level scores across identity treatment conditions. Figure 10 shows that American respondents displayed slightly higher k-level reasoning than Chinese respondents \((W = 1127300, p < 0.000)\), though the differences are small; 84.0% of American respondents were 0-level reasoners, while 89.7% of Chinese respondents were 0-level reasoners. It is therefore unlikely that the crossnational differences we obtain in the main paper are
due to differences in k-level reasoning.

Figure 10: Chinese sample not substantially different in k-level reasoning than American sample

The second potentially relevant compositional difference involves attitudes towards the use of military force, or what the public opinion about foreign policy literature often calls militant assertiveness (Herrmann, Tetlock and Visser, 1999; Kertzer and Brutger, 2016), the distinction between hawks and doves, who vary in their beliefs about the desirability and efficacy of the use of force. Militant assertiveness is worth investigating in this context given widespread debates in American foreign policy circles about the extent of the Chinese government’s new “assertiveness” in the region (Johnston, 2013), and the demonstrated importance of the construct in the study of support for the use of force (Herrmann, Tetlock and Visser, 1999). In the US sample, we measure attitudes towards the use of force with three items from the existing literature: two capturing beliefs about the efficacy of the use of force in general (“Going to war is unfortunate but sometimes the only solution to international problems”, ”The use of military force only makes problems worse”), and the third specifically implicating the role of one’s own military (“The best way to ensure peace is through American military strength”) (Kertzer and Renshon, 2014; Rathbun et al., 2016). Both are Likert items, scaled from “Strongly agree” to “Strongly disagree,” and reverse-coded where applicable such that higher answers correspond to higher levels of hawkishness. In the Chinese sample, our military assertiveness measure consists of two of these items, one capturing general beliefs about the efficacy of the use of force (“Going to war is unfortunate but sometimes the only solution to international problems”) and the other focusing specifically on the Chinese military (“The best way to ensure
peace is through Chinese military strength"). In the analysis below, we therefore operationalize atti-
titudes to the use of force both using construct-level comparisons (additive scales utilizing all of the
military assertiveness items measured in each country) and item-level comparisons (focusing solely
on items fielded in both samples).  

Figure 11: Chinese sample more likely to see military strength as necessary for peace

Figure 11 shows that the Chinese sample is significantly more hawkish than the American
counterpart ($t = -21.9, p < 0.000$); item-level comparisons suggest the difference is not attributable
to beliefs about the efficacy of force in general, but because Chinese respondents are significantly
more likely to agree that the best way to ensure peace is through Chinese military strength than
American respondents are about American military strength ($W = 622460, p < 0.000$). Yet when we
look at the marginal effects presented in Figure 12, it shows that differences in hawkishness cannot
explain the divergent results. For example, it looks as if peace through strength moderates the effect
of perspective-taking in the Chinese sample when the US escalates, but the marginal effect appears
negative, rather than positive as with American respondents.

The final potentially relevant compositional difference involves dispositional perspective taking.
Although perspective-taking is studied situationally through random assignment in the main text,
we can also study it as an individual-difference, using the dispositional measure from Davis (1980)
described in section 3 of the appendix. As with the other respondent-level variables studied above,
our empirical strategy here has two steps. First, do two samples significantly differ from one another

\footnote{As one expects, reliability estimates are higher for the three-item scale than the two-item scale: In the US sample: $\alpha = 0.69$; in the Chinese sample: $\alpha = 0.52$.}
Figure 12: Hawkishness does not significantly moderate the impact of perspective-taking

Note: each panel displays the marginal effect of the perspective-taking treatment across a different dispositional variable, estimated using a kernel estimation procedure with 5-fold least-squares cross-validation using the interflex package in R to avoid making assumptions about the functional form of the interactive effect.
on the trait in question? Second, do respondents with varying levels of the trait respond differently to the perspective-taking treatment? Figure 13 shows that overall, our American respondents are higher in self-reported perspective-taking than our Chinese respondents ($t = 8.76, p < 0.000$). Figure 14, however, shows that differences in dispositional perspective taking are unlikely to explain cross-national differences in our results. We see a strong linear interaction effect for the US sample when the other side escalates — corresponding with the results presented elsewhere in the Appendix — but it is not the case that those who are less likely to naturally engage in perspective-taking among the American respondents in the left-hand half of that panel behave more like their Chinese counterparts, given that the effect of perspective-taking amongst Chinese respondents in a similar condition is negative.

Figure 13: American sample slightly higher in dispositional perspective taking

![Dispositional empathy distribution](image)

Note: US sample in blue, Chinese sample in red. The mean of each distribution is represented by a vertical line.
Figure 14: Variation in dispositional perspective taking unlikely to explain cross-national differences

Note: each panel displays the marginal effect of the perspective-taking treatment across a different dispositional variable, estimated using a kernel estimation procedure with 5-fold least-squares cross-validation using the interflex package in R to avoid making assumptions about the functional form of the interactive effect.
6.4 Explaining individual-level variation in attribution asymmetry

We can use the same theoretical mechanisms posted above to explain individual-level variation in the attribution asymmetries we observe within each sample. As we note in the main text, our theory of perspective taking focuses on three mechanisms in particular. The first is the nature of the metasterotypes that individuals draw upon when generating attributions for the other side. Individuals who draw on metasterotypes that are based on negative or threatening images are more likely to believe the other side is acting in a threatening or offensively motivated manner. This not only increases the probability that an individual attributes another country’s actions to being offensively motivated, but also reinforces the belief that their side is acting in a defensive manner to counter the other country’s threat. This implies that two people evaluating the strategies of two countries may come to very different attributions based on the metasterotypes that come to mind when they attempt to consider the motivations of the other side. The more negative the metasterotypes that someone draws upon the greater the attribution asymmetry they are likely to perceive.

Similarly, perceived threats to one’s goals and identity will also shape how attributions are formed. If someone views another country as directly threatening to their country’s goals or identity, they are likely to view their country’s response as a defensive action to protect their interests and identity. The inverse is also true, in that the same individual is likely to view the other country’s actions as being more offensively motivated if they believe the actions are designed to threaten their interests or identity. However, different people may perceive the same situation through different knowledge structures that shape how these evaluations are made. At the country-level it is easy to recognize that different countries have different narratives about the history of conflicts and the motivations of the actors involved, which shape the knowledge structures that the country’s citizens draw upon when evaluating a new situation. At an individual-level, we know that perceptions of threat to one’s goals and identity are shaped by a multitude of factors that vary both within and across countries. Thus, individuals who perceive another country as being more threatening are likely to view that country’s actions as being more offensively motivated, whereas individuals who do not perceive another country as a threat are likely to view that country’s actions as being less aggressive and more defensively motivated.

In the context of our study and the US-China relationship, we can draw a few empirical im-
lications from this theoretical framework. First, we expect that those individuals who are highly nationalistic, and are thus more concerned with threats to their national identity, are likely to exhibit larger attribution asymmetries. We further expect that those who perceive the other country as being a greater threat — either because of their negative intentions, or because of the greater power they possess — are also likely to exhibit greater attribution asymmetries.

We begin our empirical examination by comparing threat perceptions, relative power parity, and nationalism across the American and Chinese samples. Beginning with nationalism, we find that Chinese respondents are significantly more nationalist than our American respondents are ($t = 3.83$, $p < 0.001$). Given Chinese respondents’ heightened sense of nationalism, our theory predicts that they should display larger attribution asymmetries, given the perceived threat to their identity. We next evaluate whether the American and Chinese respondents have different threat perceptions in regards to one another. Again, we find that the Chinese respondents view the US as having significantly more threatening goals than Americans view China (difference in mean: 0.605, $p < 0.001$), which should also lead the Chinese to display a greater attribution asymmetry. Lastly, when asked about the relative power of the other country compared to their own, Chinese respondents believe the US is significantly more powerful than China, whereas American respondents are more likely to see the two as being relative equals and near parity (difference in mean: 1.268, $p < 0.001$). This implies that Americans, who view the two countries have being closer to parity, should have lower attribution asymmetries. Linking the last two differences shows that the Chinese not only believe the US has more threatening intentions, but they believe the US is stronger and thus has more power to pursue its threatening goals. All of these indicators point to the Chinese being more likely to view US actions as aggressive and offensively motivated, while also giving them reason to believe their own actions are defensively motivated to counter the American threat.

We next examine the relationship between attribution asymmetries and threat perceptions, relative power parity, and nationalism in each sample, as is shown in Table 6. Consistent with our theory, we find that increased threat perceptions and higher levels of nationalism are generally associated with greater attribution asymmetries. We also find that the perception of relative power parity, as opposed to believing the other side is stronger, is associated with lower attribution asymmetries. Given these relationships and the systematic differences in nationalism, threat perception, and perceived power parity across the US and Chinese samples, we now have evidence demonstrating that the differing knowledge constructs, both within and across countries, contribute to the difference in
magnitude of attribution asymmetries between the US and China.

Table 6: Attribution Asymmetry, Threat Perception, and Nationalism

<table>
<thead>
<tr>
<th></th>
<th>US Sample</th>
<th></th>
<th>China Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Threat perception</td>
<td>0.417***</td>
<td>0.304***</td>
<td>0.378***</td>
<td>0.381***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.045)</td>
<td>(0.037)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Power parity</td>
<td>-0.146***</td>
<td>-0.099**</td>
<td>-0.240***</td>
<td>-0.274***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.044)</td>
<td>(0.042)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>National attachment</td>
<td>0.705**</td>
<td>-0.056</td>
<td>0.192***</td>
<td>0.193***</td>
</tr>
<tr>
<td></td>
<td>(0.315)</td>
<td>(0.324)</td>
<td>(0.033)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Nationalism</td>
<td></td>
<td></td>
<td>0.192***</td>
<td>0.193***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.033)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Military assertiveness</td>
<td>2.281***</td>
<td></td>
<td>0.339</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.345)</td>
<td></td>
<td>(0.245)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.667***</td>
<td></td>
<td>-1.557***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.248)</td>
<td></td>
<td>(0.420)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.014</td>
<td></td>
<td>-0.321***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.124)</td>
<td></td>
<td>(0.123)</td>
<td></td>
</tr>
<tr>
<td>Ideology</td>
<td>0.549**</td>
<td></td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.248)</td>
<td></td>
<td>(0.134)</td>
<td></td>
</tr>
<tr>
<td>Party Member</td>
<td></td>
<td></td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.134)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.746***</td>
<td>-1.872***</td>
<td>0.104</td>
<td>1.063***</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.303)</td>
<td>(0.305)</td>
<td>(0.406)</td>
</tr>
<tr>
<td>N</td>
<td>1.492</td>
<td>1.462</td>
<td>1.479</td>
<td>1.465</td>
</tr>
<tr>
<td>R²</td>
<td>0.071</td>
<td>0.111</td>
<td>0.108</td>
<td>0.120</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.069</td>
<td>0.107</td>
<td>0.106</td>
<td>0.116</td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01
References


