
Hawkish Biases and Group Decision Making

Joshua D. Kertzer,  Marcus Holmes,  Brad L. LeVeck, 
and Carly Wayne 

Abstract How do cognitive biases relevant to foreign policy decision making aggregate in groups? Many tendencies identified in the behavioral decision-making literature—such as reactive devaluation, the intentionality bias, and risk seeking in the domain of losses—have been linked to hawkishness in foreign policy choices, potentially increasing the risk of conflict, but how these “hawkish biases” operate in the small-group contexts in which foreign policy decisions are often made is unknown. We field three large-scale group experiments to test how these biases aggregate in groups. We find that groups are just as susceptible as individuals to these canonical biases, with neither hierarchical nor horizontal group decision-making structures significantly attenuating the magnitude of bias. Moreover, diverse groups perform similarly to more homogeneous ones, exhibiting similar degrees of bias and marginally increased risk of dissension. These results suggest that at least with these types of biases, the “aggregation problem” may be less problematic for psychological theories in international relations than some critics have argued. This has important implications for understanding foreign policy decision making, the role of group processes, and the behavioral revolution in international relations.

The past several decades have seen a surge of interest in psychological approaches to the study of international politics.¹ Unlike structural realist or rationalist approaches, which largely study features of the environments in which actors are embedded, psychological theories of international politics turn to the properties of actors themselves.² A large volume of literature has thus emerged on the psychology of political elites: their operational codes, personality traits and leadership styles, and so on.³ One of the central insights of this literature is that leaders are imbued with many of the same psychological mechanisms as ordinary citizens: they are prone to misperceptions, engage in motivated reasoning, and rely on heuristics and biases.⁴

1. For a review, see Davis and McDermott 2021; Hafner-Burton et al. 2017; Kertzer and Tingley 2018; Levy 2013; Mintz 2007.

2. Holmes 2018; Kertzer 2016; Lake and Powell 1999; Landau-Wells 2018; Larson 1985; McDermott 1998; Powers 2022; Rathbun 2014; Renshon 2017; Saunders 2011; Vertzberger 1998; Waltz 1979; Yarhi-Milo 2018.

3. Etheredge 1978; George 1969; Greenstein 1969; Hermann 1980; Leites 1951.

4. Baekgaard et al. 2019; Brooks, Cunha, and Mosley 2015; Jervis 1976; D. Johnson 2020; Kertzer 2021; Kertzer, Rathbun, and Rathbun 2020; LeVeck et al. 2014; Poulsen and Aisbett 2013; Sheffer et al. 2018; Stein 1988.

The presence of these biases in decision making is of particular importance. As Kahneman and Renshon note, in the context of foreign policy, nearly all of the cognitive biases uncovered by psychologists would lead political leaders to make more *hawkish* decisions, all else equal.⁵ That is, these tendencies increase suspicion, hostility, and aggression toward potential adversaries, increasing the risk of political conflict and violence.⁶ Individuals' tendency to take risks to avoid a loss, for example, could encourage leaders to prolong wars beyond the point at which victory is achievable, engaging in risky offensives with little chance of success.⁷ Likewise, leaders may become less willing to make concessions and more willing to risk large losses when bargaining.⁸ The biased ways in which people assess the motives of adversaries could also increase the potential for conflict.⁹ For instance, individuals tend to assess the intentionality of an act by its *consequences*, rather than by a thorough examination of the perpetrator's motives.¹⁰ As a result, wartime actions that produce morally bad outcomes are more likely to be deemed intentional than identical actions that produce morally good outcomes.¹¹ Yet another cognitive bias that can prolong or worsen conflict is reactive devaluation, the tendency of individuals to immediately discount or devalue proposals coming from an adversary, compared to identical proposals offered by one's own side or a third-party mediator.¹²

Yet for all of its rich insights, this literature has wrestled with a challenge. Most of what scholars know about psychological biases in decision making comes from the study of *individuals*, but many foreign policy decisions are made in *group* contexts. Indeed, groups are often used in foreign policy decision-making settings precisely because of their (presumed) ability to counter the decision-making pathologies or shortcomings of individuals acting in isolation.¹³ Thus the theoretical and empirical value of insights from the behavioral sciences on the pathologies of individual decision making are often criticized in the study of foreign policy for a lack of clear understanding of how preferences, information, or traits aggregate into group-level decisions, with critics typically arguing that these psychological biases should be

5. Kahneman and Renshon 2007. See also D. Johnson 2020, 268. While our interest here is on three biases that tend to move in a hawkish direction with respect to decision making, it may be that others have a tendency to move in a dovish direction, or create misperceptions that lead to cooperation rather than the use of force. See Grynaviski 2014.

6. Kahneman and Renshon 2009. We follow Kahneman and Renshon 2007 in referring to these phenomena as "hawkish biases," but we do not use the term in a pejorative sense, or to imply that these tendencies are inherently irrational—see, for example, Gigerenzer and Gaissmaier 2011; D. Johnson 2020. We can think of these tendencies more generally as what behavioral scientists refer to as "nonstandard" preferences, beliefs, and decision making, behavioral regularities traditionally excluded from canonical rational choice models, as in DellaVigna 2009; Hafner-Burton et al. 2017. For an application of hawkishness to international relations more generally, see Mattes and Weeks 2019.

7. Kahneman and Tversky 1979; McDermott 1998.

8. Levy 1996.

9. Jervis 1976.

10. Knobe 2003.

11. Chu, Holmes, and Traven 2021.

12. Ashmore et al. 1979; Maoz et al. 2002; Ross and Ward 1995.

13. Hart, Stern, and Sundelius 1997.

mitigated or otherwise cancel out in group settings.¹⁴ Even proponents of psychological approaches have noted this limitation. In an important review of prospect theory, for example, Levy notes that “Most of what we want to explain in international politics involves the actions and interactions of states ... each of which is, in principle, a collective decision-making body. The concepts of loss aversion, the reflection of risk orientations, and framing were developed for individual decision making and tested on individuals, not on groups, and we cannot automatically assume that these concepts and hypotheses apply equally well at the collective level.”¹⁵ Writing two decades later, Hafner-Burton and colleagues express a similar concern, noting that institutional structures are often designed precisely to mitigate individual psychological biases.¹⁶

Ultimately, however, the question of how psychological biases in foreign policy aggregate in groups—and whether groups indeed attenuate these biases—remains an empirical one, as theories of aggregation provide few guarantees. For example, Arrow’s famous “impossibility theorem” shows that, even if all the individuals in a group are perfectly rational and calculating, many aggregation mechanisms can still produce irrational choices.¹⁷ Meanwhile, other theorems show that aggregation can lead to more optimal decision making. However, such improvement often requires a set of fairly restrictive assumptions. For example, Condorcet’s well-known jury theorem shows that sufficiently large groups can make better decisions if each individual votes independently and makes the right choice with probability greater than 50 percent. Yet, violating any of these assumptions may actually cause groups to make *worse* decisions than individuals.¹⁸ This could be particularly concerning in many foreign policy decision-making contexts, where policy is often decided by small groups of individuals who influence one another and who may be systematically biased toward the wrong decision.¹⁹

In this piece, we offer what we believe to be the first direct experimental test of the aggregation of psychological biases in foreign policy. We field three large-scale online experiments, where nearly 4,000 participants work through a series of foreign policy scenarios, which they completed either as individuals, or in one of two different types of group structures. We find that three prominent tendencies from the behavioral decision-making literature—risk taking to avoid a loss, the intentionality bias, and reactive devaluation—largely replicate in small-group contexts. We find no evidence that these tendencies are significantly reduced in group settings, and find that in some decision-making contexts they may even be exacerbated. Moreover, we find little evidence that more experienced leaders can improve group decision making or that more diverse groups are less prone to hawkish

14. Powell 2017; Saunders 2017, S220.

15. Levy 1997, 102.

16. Hafner-Burton et al. 2017, S18–S21.

17. Arrow 1950.

18. Austen-Smith and Banks 1996.

19. Janis 1972.

biases. These findings have important implications for how we understand the role of group processes in foreign policymaking, suggesting that groups are not a panacea for producing optimal policy decisions, and that we should not assume that the psychological tendencies that shape individual decision making do not appear in collective contexts as well.

Biases and Group Decision Making

The question of how group processes affect decision making is not a new one. Indeed, outside of international politics, there is a rich and diverse literature that has explored the ways in which group settings affect bias and judgment. In legal studies, for example, research on jury decision making explores how juror-level characteristics aggregate in shaping jury-level decisions.²⁰ In business administration, organizational behavior research focuses on how the traits of team members have varying effects on team performance depending on the types of tasks.²¹ In social network analysis, scholars have experimentally studied the conditions under which collective decision making outperforms individual decision making.²² Indeed, a small cottage industry has now formed that includes interdisciplinary approaches to “small group decision making,” which investigates, among other things, individual cognitive biases and under what conditions they might be overcome (or exacerbated) in a group setting. Even nonhuman animal models might offer relevant insights. A school of fish can follow light too weak for any individual fish to follow, for example.²³

While this diverse scholarship may offer crucial insights for the study of foreign policy, it has important limitations. Many invocations of the “aggregation problem” in political science are more philosophical than empirical, assuming *ex ante* that aggregation is a challenge rather than empirically testing the specific contexts in which psychological variables should or should not aggregate.²⁴ Because of the high cost of bringing large numbers of people into the lab, many of the canonical experimental tests of aggregation in group decision making have traditionally been somewhat underpowered, testing the impact of relatively small groups.²⁵ Thus it has been difficult to identify what aspects of group decision making causally affect outcomes. Perhaps most importantly, foreign policy decision making involves three theoretically relevant institutional structures and task properties that differentiate it from some of the main configurations frequently studied in the literature outside political science.

20. Devine et al. 2001.

21. Moynihan and Peterson 2001.

22. Bernstein, Shore, and Lazer 2018.

23. Berdahl et al. 2013.

24. See, for example, Gildea 2020; D. Johnson 2015; Mercer 1995, 237–38; Powell 2017; Wendt 2004.

25. E.g., Lewin, Lippitt, and White 1939.

First, foreign policy decision making, particularly over security issues, often features ill-structured problems, where the probability distributions may be unknown.²⁶ Actors may not know, or may disagree on, the parameters of the decision-making task; they may even disagree on the ultimate goal with respect to the decision to be made. These situations stand in contrast to much, though not all, of the small-group research and analysis of aggregation that occur in other disciplines. Investigations of cognitive biases, for example, often use well-structured problems with clear probability distributions. Alternatively, studies that investigate the “wisdom of crowds” will often use difficult, but nevertheless clearly structured, math problems.²⁷ It therefore remains unclear how generalizable insights from clearly structured problems may be to decision making in the more amorphous context that characterizes much of international politics.

Second, foreign policy decision making often involves hierarchically structured groups, where the chain of command and the decision-making rules are known to all the actors involved. While the existing research on small group dynamics and decision making in groups takes many forms, including analysis of groups within large-scale hierarchical settings such as firms, much of the research political science has brought in has tended to focus on “flat” or horizontal groups, such as teams, and has not systematically compared the effects of hierarchical versus horizontal decision-making structures.²⁸ Hierarchies may emerge endogenously over time as a result of specific group members’ personalities, but this is theoretically very different from ingrained hierarchies built on formal and clear roles and decision-making rules.²⁹ It is partly because of the hierarchical nature of many foreign policy institutions that much of the foreign policy decision-making literature focuses on leaders, rather than advisers.³⁰ Moreover, without manipulating these structural conditions it is difficult to gain analytical leverage on how hierarchy affects foreign policy decision making.

Third, the substantive focus of scholars of foreign policy decision making, including distinctive outcomes of interest, are often very different from those studied in small-group research in other domains. Analysts of foreign policy are often interested in explaining specific dependent variables, such as a decision to use force. These are quite different from those often studied in small-group research, such as team morale or workplace satisfaction in a business context, or performance on mathematical exercises. It may be that the specific decisions of interest, such as the use of force, engage different aggregation processes, limiting the utility of extrapolating findings from small-group research to foreign policy.

Empirical research in political science has tended to focus on how groups might *improve* decision making, which brings in a normative component, and has returned

26. Brutger and Kertzer 2018; Voss and Post 1988.

27. LeVeck and Narang 2017.

28. Kerr and Tindale 2004; Larrick 2016; LeVeck and Narang 2017.

29. Strodbeck, James, and Hawkins 1957.

30. Though see Ausderan 2013; Kaarbo 1998; Redd 2002; Saunders 2017; Weeks 2014.

a mixed bag of results: factors such as group size, composition, decision-making rules, political context, and leadership can all affect the quality of the decision-making process and outcome.³¹ For example, groupthink, the most famous psychological dynamic documented in political group decision making, whereby group members' striving for unanimity exacerbates decision-making pathologies, is hypothesized to be a *contingent* phenomenon, most likely to emerge under conditions of strong social-unit cohesion and external stress.³²

Driven by this finding, as well as subsequent research affirming the danger of group members' striving for unanimity, many of the most prominent proposals for improving the quality of foreign policy decision making focus on constructing a diverse decision unit, led by an experienced leader who fosters healthy debate and dissent in the policymaking process.³³ These principles guide decision-making models such as multiple advocacy, the competitive advisory system, and distributed decision making.³⁴ Indeed, the perceived value of diversity as a tool to harness the mental power of groups and improve decision making is a hallmark of much recent scholarship.³⁵ However, diversity is not without risk, and may also increase intragroup conflict and decision paralysis.³⁶ Thus the benefits of diversity in improving decision making may depend on the presence of a leader who is well positioned to channel that diversity in productive directions. For example, research has suggested that a leader's experience, leadership style, predispositions, and personality can all shape their ability to harness the information-processing power of groups to improve decision making.³⁷ However, most research in political science on group decision making has relied on small-*N* case studies, which limits our ability to identify how different attributes of the group setting, such as the distribution of information individuals have or the experience they bring to the table, affect the quality of decision making.

In sum, while there are impressive cognate bodies of literature on aggregation outside of political science, and rich descriptive evidence on group dynamics in policymaking settings, we do not yet have strong experimental evidence regarding the effects of groups in the complex settings that characterize foreign policy decision making, nor do we fully understand how different decision rules, group composition, and leader attributes shape these processes.

In this study we test for the effects of group decision making on the prevalence of three well-known cognitive biases that have been observed in individual decision making: risk taking to avoid a loss, the intentionality bias, and reactive devaluation.³⁸

31. Kerr, MacCoun, and Kramer 1996.

32. Hart, Stern, and Sundelius 1997; Janis 1972.

33. Esser 1998; Sunstein and Hastie 2014.

34. George 1972; R.T. Johnson 1974; Schneeweiss 2012.

35. Horowitz et al. 2019; Page 2019.

36. Mintz and Wayne 2016.

37. Herrmann and Preston 1994; Horowitz and Fuhrmann 2018; Preston 2001; Saunders 2017; Schafer and Crichlow 2010.

38. Kahneman and Tversky 1979; Knobe 2003; Ross and Ward 1995.

Each of these biases has been theorized to bias political elites in a “hawkish” direction.³⁹ In other words, all else equal, the presence of these biases may cause leaders to demonstrate a greater “propensity for suspicion, hostility, and aggression in the conduct of conflict, and for less cooperation and trust when the resolution of conflict is on the agenda” than is objectively warranted.⁴⁰

For example, loss aversion could reduce leaders’ willingness to compromise in negotiations. Their own concessions would be viewed as “losses,” while an adversary’s concessions would be viewed as “gains”—and even when these concessions are equal, the gains would *feel* smaller than the losses, and so compromises would likely be rejected.⁴¹ Similarly, the intentionality bias, whereby individuals assess whether an action was intentional based on its effects, may lead to misperceptions or unfounded certainty regarding intentionality. Actions with negative consequences, or “side effects,” are more likely to be seen as intentional. Such ascriptions are relevant in a range of contexts, from security dilemma escalation to public assessments of blame in civil conflicts.⁴² Finally, reactive devaluation—a bias whereby a proposal is automatically perceived as less valuable if offered by an adversary—has been shown to affect attitudes toward negotiations in various political conflicts, from US–Soviet interactions during the Cold War to the ongoing Israeli–Palestinian conflict.⁴³ Together, then, these three biases have the potential to reduce the likelihood of negotiation success and trigger or prolong violent political conflict. Assessing the extent to which these individual-level biases scale to affect foreign policy decisions that are often made in group contexts is crucial for understanding how the institutional structures of foreign policymaking potentially mitigate or exacerbate the influence of these biases on international cooperation and conflict.

Research Design

The present study aims to examine the relative efficacy of groups in reducing the impact of these biases on decision making using three large-scale online group experiments conducted in Fall 2019 and Winter 2020, whose structure is summarized in [Figure 1](#).⁴⁴ By manipulating the group setting, this study provides causal leverage to examine how the cognitive biases of individuals aggregate in different types of group decision-making units. As with all experiments, there are important questions about external validity to keep in mind, which we discuss in detail later.

39. Kahneman and Renshon 2007.

40. Kahneman and Renshon 2009, 79.

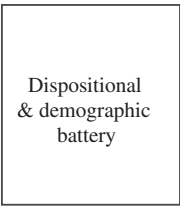
41. Kahneman and Tversky 2017.

42. Mitzen and Schweller 2011; Pechenkina and Argo 2020.

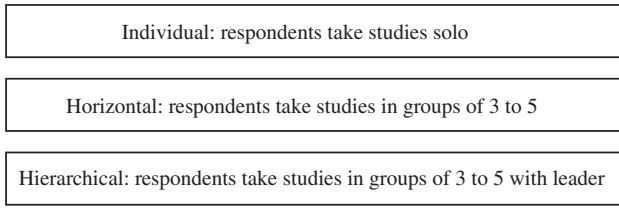
43. Ashmore et al. 1979; Maoz et al. 2002.

44. Respondents were a sample of adults in the United States recruited using Qualtrics. Qualtrics is a panel aggregator, so it has access to a much larger sample than any single online panel, which is necessary to produce a sufficient flow of respondents for successful synchronous group interaction.

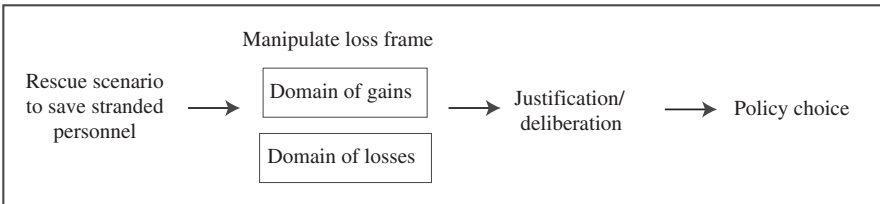
1. Demographics



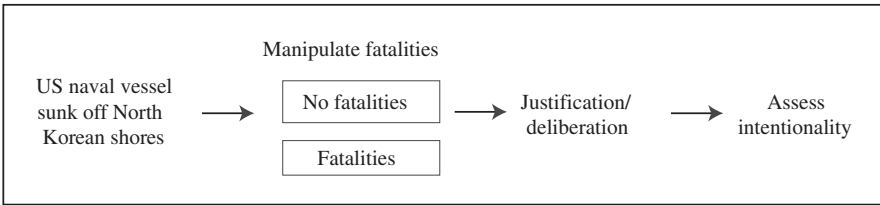
2. Group assignment



3. Prospect theory experiment



4. Intentionality bias experiment



5. Reactive devaluation experiment

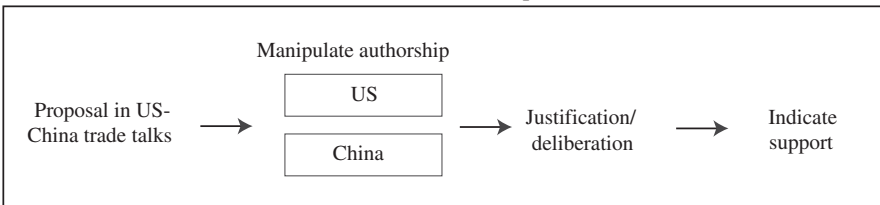


FIGURE 1. Study design

The study proceeds as follows. After completing an individual-differences and demographic battery, respondents are randomly assigned to one of three group conditions. In the individual condition, 760 respondents are asked to make decisions on various foreign policy scenarios *individually*, taking notes as they think through their options. In the two group conditions, respondents are assigned to a group with four other survey takers, in which they participate in a group chatroom, discussing their

options together before deciding on a course of action. There are two types of groups: *horizontal* groups, where participants are asked to try to come to a collective, unanimous decision and each participant has equal say in the process; and *hierarchical* groups, in which one of the five participants is randomly designated as the leader of the group, and gets to make the final choice, in consultation with the four other participants, who take on the role of adviser. In the analysis that follows, the group conditions consist of 3,213 respondents, forming 771 groups (406 horizontal, 365 hierarchical) of up to five members each. We paid an average of USD 10 per subject in respondent incentives, and all together, the effective sample size (N) of the study is 3,987.⁴⁵

After being assigned to one of these treatments, respondents pass through three separate experimental modules using canonical experimental setups to examine the prevalence of various biases in the context of foreign policy decision-making scenarios. Respondents in the individual condition complete these modules as individuals, writing down their justifications for their decisions and making decisions themselves, whereas respondents in the group conditions complete these modules as groups, deliberating as a group before reaching decisions.⁴⁶ An example of a group deliberation is shown in Figure 2. Respondents were generally engaged in the group deliberations; in the horizontal condition, 73 to 76 percent of group members in the analysis participated more than once in each deliberation, similar to the rate observed in the hierarchical condition (74 to 81 percent), with leaders participating more frequently than advisers—though as we show in section 4 of the online supplement, our findings are robust and do not significantly vary across different levels of group participation.

The first experimental module examines sensitivity to gain and loss frames on policy preferences—a canonical finding from prospect theory. Subjects are presented with a scenario in which “600 lives are at stake in a war-torn region.” Subjects are asked to choose one of two courses of action (Policy A or Policy B). Policy A will definitively lead to 200 people dying and 400 people being saved. Policy B has a probabilistic outcome, with a 1/3 probability that no one will die (all 600 will be saved) and a 2/3 probability that 600 people will die (none will be saved). The experimental treatment within this module is whether the results of each policy are presented in the domain of *gains* (e.g., “200 people will be saved”) versus the domain of *loss* (e.g., “400 people will die”). Half of the respondents in each experimental

45. These groups of five—as well as the assigned leader in hierarchical groups—stay the same throughout each of three experimental modules. That is, group members do not change from module to module, though some groups do become smaller due to dropouts; our analysis includes only groups with no fewer than three members in a given experiment; in the hierarchical condition the group must also include a leader. We also manually screened the respondents for “bots,” removing from the analysis any individual (or group, in the group conditions) that displayed bot-like behavior in the chat logs. For a detailed set of attrition tests and sensitivity analyses that show the robustness of the findings, see section 2.2 in the online supplement.

46. Respondents in the group conditions deliberated using a chat platform constructed in SMARTRIQS. See Molnar 2019.

*** Advisor1 has joined the chat ***
*** Advisor4 has joined the chat ***
*** Advisor3 has joined the chat ***
*** Leader has joined the chat ***
*** Advisor2 has joined the chat ***
Advisor3: Everyone got rescued, but by whom?
Advisor1: I'm inclined to believe it is unlikely that the ship was attacked as all survived.
Leader: Given the past erratic behavior of North koreas leaders I would say it is extremely likely
Advisor3: If NK attacked, they wouldn't have rescued anyone.
Advisor3: Unless it was an accident, and they felt bad.
Advisor4: no one said they rescued the crew
Advisor2: We should understand the details of the situation before coming to any conclusion. It is unlikely that this was a provocation. It may be an accident
Advisor1: An attack to me would mean at least a few would not have survived
Advisor4: I would assume its an unfortunate accident. assuming nk did it will cause more panic. unless another incident occur
Advisor3: It could be purely coincidental that the ship sank in the location it did. I still want to know who did the search and rescue.
Advisor2: All service people survived. So is is unlikely that this was an attack
Leader: I agree advisor 2, we need more details to come to a informed decision
Advisor4: we can amummer somewhat unlikely and just monitor any suspicious activities
Advisor1: search & rescue is moot. The question is NK attacked it or didn't
Advisor3: It would be best to not accuse NK of anything until a reason for sinking is determined.
Advisor2: I think this need more detailed investigation
Advisor3: Accusations could escalate quickly given their leader's nature.
*** Advisor1 has left the chat ***
Advisor3: I say it is unlikely.
*** Advisor4 has left the chat ***
*** Advisor3 has left the chat ***
*** Leader has left the chat ***

Notes: Transcript of a group deliberation session from one of the hierarchical groups in the intentionality bias experiment. Note that one of the group members points to the absence of fatalities as a sign the act was unintentional, consistent with the logic of intentionality bias.

FIGURE 2. *Sample group deliberation transcript*

condition (individual, horizontal group, or hierarchical group) receive the “gains” treatment and half receive the “loss” treatment.⁴⁷

The second experimental module tests susceptibility to the intentionality bias—the degree to which assessments of intentionality are affected by the (negative) results of an event. In this module, respondents are asked to assess how likely it is that a US navy vessel sunk 100 miles off the coast of North Korea was intentionally versus accidentally targeted by the North Koreans. The randomly assigned treatment in this module is the number of casualties the sinking of this vessel has caused: none versus all 100 servicepeople on board. Half of the respondents in each experimental condition receive each treatment. This represents a more ill-structured problem than that posed by the previous experiment.

The final experimental module explores the prevalence of reactive devaluation of a trade negotiations proposal between the United States and China. Subjects view a short proposal that purports to resolve ongoing US–Chinese disputes over trade. The experimental treatment is the authorship of the text—whether the United States or China drafted the proposal. As with the first two modules, half the respondents in each experimental condition receive each treatment. Instrumentation for each of the three experiments is shown in section 1 of the online supplement.

We calculate our dependent variable differently in the three modules based on the group condition. In the individual conditions, we focus on the choice of each individual respondent. In the hierarchical conditions, we focus on the choice of each group leader. In the horizontal conditions, we primarily use a median voter rule to calculate each group’s decision, but we also use two other aggregation rules (majority vote and unanimity) to test how sensitive our findings are to other means of aggregating group members’ votes. We describe these different aggregation methods in detail in section 2.1 of the online supplement.

Together, these studies are useful because they allow us to examine the extent to which hawkish biases replicate in individual settings and the degree to which group discussion—and the structure and composition of those groups—affect their prevalence, in experiments that differ from one another in a variety of ways. The existing literature lends us strong theoretical expectations in regard to the individual condition, given the canonical nature of these cognitive biases: we expect that individuals will be more risk seeking in the domain of losses than the domain of gains, will be more likely to assess an incident as intentional when its costs are higher, and will evaluate a proposal from an adversary more negatively than the same proposal from their own side.

Yet given both the novelty of our particular study and the contradictory arguments in the literature on the efficacy of groups in reducing biases, the ultimate effects of groups on these hawkish biases remains an open question. Groups could reduce the prevalence of hawkish biases, exacerbate them, or have no effect—particularly

47. All members of a single group receive the same treatment. For example, the five members of a horizontal group that have been randomly grouped together would all receive only the “gains” frame.

given that these hawkish biases may be deeply ingrained, or outside the realm of conscious awareness.⁴⁸ Empirically adjudicating between these competing expectations constitutes one of the central contributions of our study.

Analysis

To test these competing expectations, we turn to each of our three experiments in sequence. For each experiment, we first look within each group condition (individual, horizontal, hierarchical) to examine the prevalence of the hawkish bias tested (susceptibility to gains/loss framing, the intentionality bias, or reactive devaluation). We then compare these differences *across* groups to assess the extent to which these different decision-making structures affect susceptibility to each of the tested biases. Finally, we probe the robustness of our findings, assessing the degree to which various types of leader characteristics or aspects of group diversity affect susceptibility to biases and the ability to reach a decision in the first place.

Susceptibility to Gains/Loss Framing

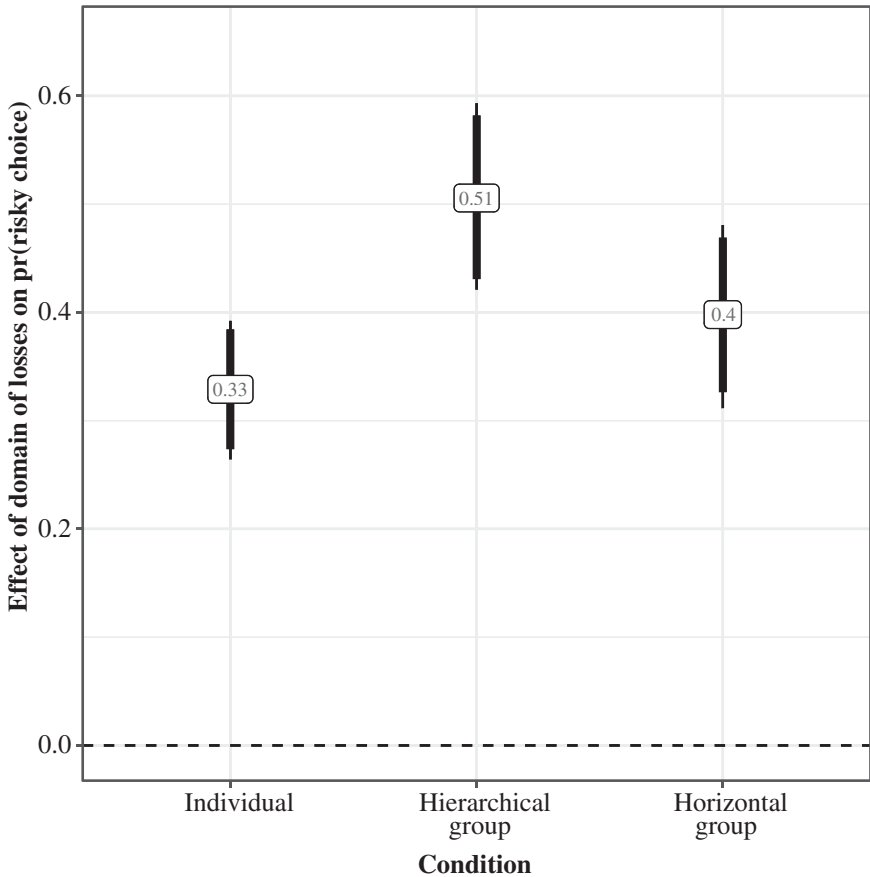
We begin by examining the prevalence of a canonical hawkish bias across our three group formulations: the effects of loss-versus-gains framing on individuals' acceptance or avoidance of risky choice.

In the individual condition, our results strongly replicate the core finding of prospect theory. When choices are framed as a potential loss (e.g., of life) individuals are significantly more likely to choose the probabilistic policy—that is, they are more accepting of the risk that all 600 lives will be lost, in order to preserve the possibility of an outcome where no one dies. In contrast, those presented with a gains framework, where people may be *saved*, are much more risk *averse*, preferring the nonprobabilistic Policy A (200 people will be saved).

Do groups reduce susceptibility to this bias? Our results suggest they do *not*; if anything, groups may increase the effect of frames on choice. In both types of groups, groups randomly presented with loss frames are significantly more likely to prefer the probabilistic outcome than groups that were presented with a gain frame (Figure 3). Examining the magnitude of these effect sizes across decision-making structures, we find that hierarchical groups in particular are significantly more sensitive than individuals to framing effects.⁴⁹

48. D. Johnson 2020; Myers and Lamm 1976; Powell 2017.

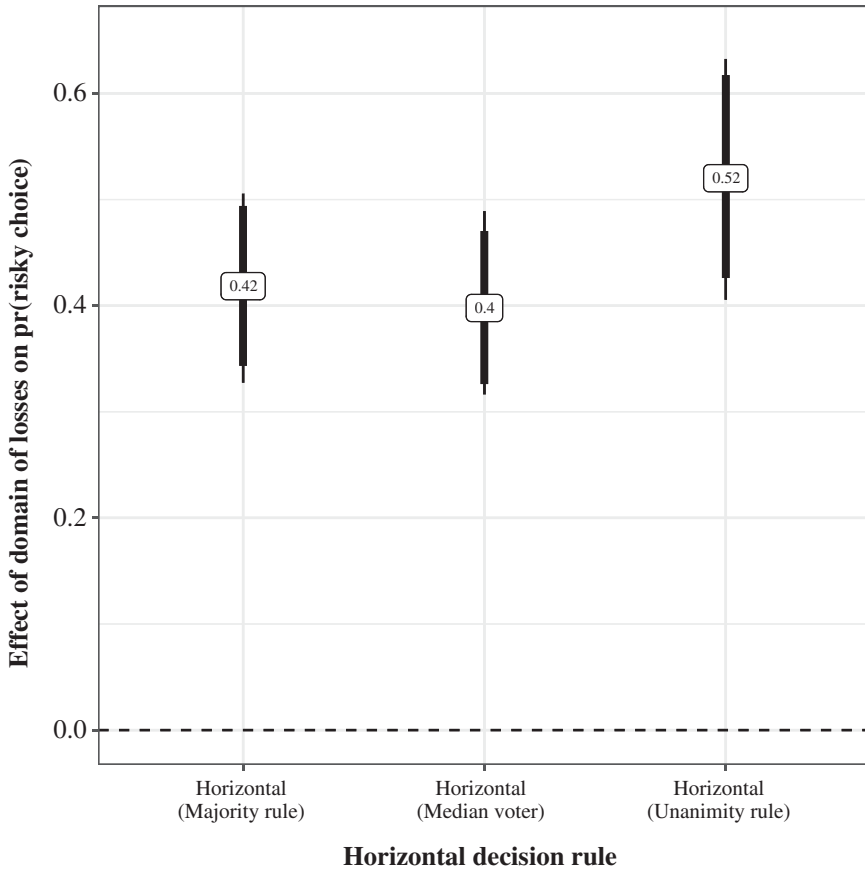
49. When comparing across groups, we use a variety of methods to account for potential covariate imbalance between individual and group conditions. Results are substantively similar regardless. Without controls, $p < .002$; with a series of controls for leader-level characteristics, $p < .003$; and with group-level controls (demographic characteristics averaged across all group members), $p < .002$ (see section 2.1 of the online supplement).



Notes: The figure shows the effect of the domain of losses on the probability of risky choice, within each group context (individual, hierarchical, or horizontal). The canonical prospect theory result is seen in both hierarchical and horizontal groups, and is exacerbated in hierarchical groups. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals. Horizontal group decisions are calculated here using the median voter aggregation method. See Figure 4 for additional aggregation method results.

FIGURE 3. *Prospect theory framing effects replicate in groups*

Comparing the horizontal groups to individual decision makers, Figure 4 suggests that the susceptibility to gain/loss frames may depend on the specific decision rule used to assess these groups. For example, examining horizontal groups that succeeded in reaching a *unanimous* decision, we find similar results as in the hierarchical condition: the group setting *increases* susceptibility to these framing effects ($p < .005$). However, if we examine the full set of horizontal groups using a less stringent decision rule, such as a majority rule ($p < .09$) or median voter ($p < .16$), we do



Notes: The figure shows the effect of the domain of losses on the probability of risky choice, within horizontal groups, using different aggregation methods. The canonical prospect theory result is seen across all three types of horizontal aggregation method (majority rule, median voter, and unanimity rule), but is the largest in unanimous groups (significantly larger than in the individual condition, $p < .005$). Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals.

FIGURE 4. Prospect theory framing effects by horizontal aggregation method

not find evidence that horizontal groups perform significantly differently than individuals. Either way, it is clear that horizontal groups do not reduce susceptibility to prospect theory’s framing effects.

Intentionality Bias

Next, we examine the relative prevalence of the intentionality bias across group settings. While the prospect theory module examines a fairly well-defined decision

problem where each policy choice features known probability outcomes, the intentionality bias module examines a more complex choice: how likely do you think it is that an event was caused by a purposeful attack by an adversary? In the individual condition, our results again strongly replicate the canonical intentionality bias finding. When the consequences of an event are more negative (in this case causing fatalities), individuals are significantly more likely to assess the event as an intentional provocation rather than the result of an accident or miscommunication. Group settings do little to attenuate this tendency: both horizontal and hierarchical groups are significantly more likely to assess the sinking of a US navy ship as the consequence of an intentional attack by the North Koreans when there are fatalities reported (Figure 5).

However, unlike the prospect theory experiment, with the intentionality bias, we find that groups have no effect on the severity of this tendency. While certain group configurations tended to make our respondents somewhat *more* susceptible to framing effects, in this case groups perform similarly to individuals—no better or worse.⁵⁰

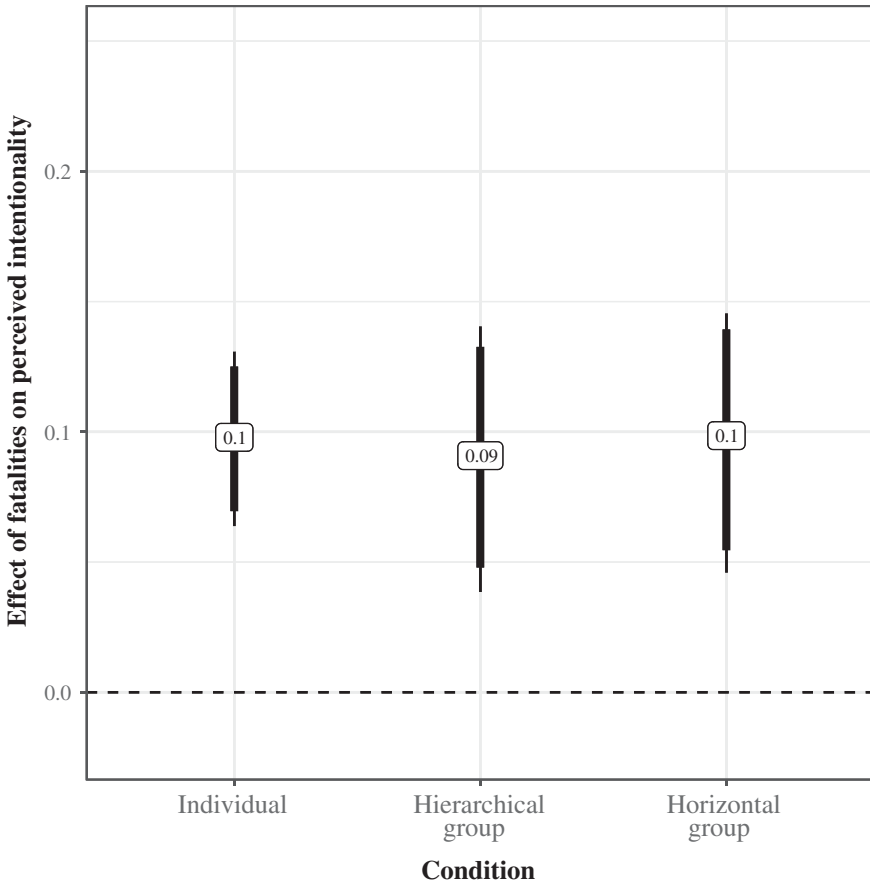
As before, horizontal groups that reach a *unanimous* decision do display a somewhat more pronounced bias than those assessed with less stringent decision rules (majority rule or median voter), but this difference is not statistically significant (Figure 6). Regardless of the aggregation method, both horizontal and hierarchical groups increase their assessments of intentionality in response to negative outcomes to a similar extent as individuals do.

Reactive Devaluation

Finally, we turn to reactive devaluation. Here we unexpectedly do *not* replicate the standard reactive devaluation result in two of the three decision-making conditions (Figure 7). Individuals are *not* significantly less likely to support a proposal authored by China than one authored by the United States. Hierarchical groups, where the decision is ultimately made by a single individual after group discussion, also do not prefer US-authored proposals.

On the one hand, this finding is surprising: the theoretical expectation is that proposals written by an adversary (e.g., China) will be automatically devalued with respect to proposals written by one's own side (the United States). However, work on reactive devaluation also suggests that there are two distinct mechanisms by which proposals are devalued: reactance processes that lead individuals to devalue that which is available compared to what is not, and reliance on source credibility

50. Comparing across groups, the difference in the effect of fatalities between the individual and horizontal condition (using the median voter rule) has $p < .94$ without controls, $p < .91$ with controls. The difference in the effect of fatalities on assessments of intentionality between the individual and hierarchical condition has $p < .86$ without controls, $p < .85$ with controls at the leader level, and $p < .85$ with controls at the group level (see section 2.1 in the online supplement).

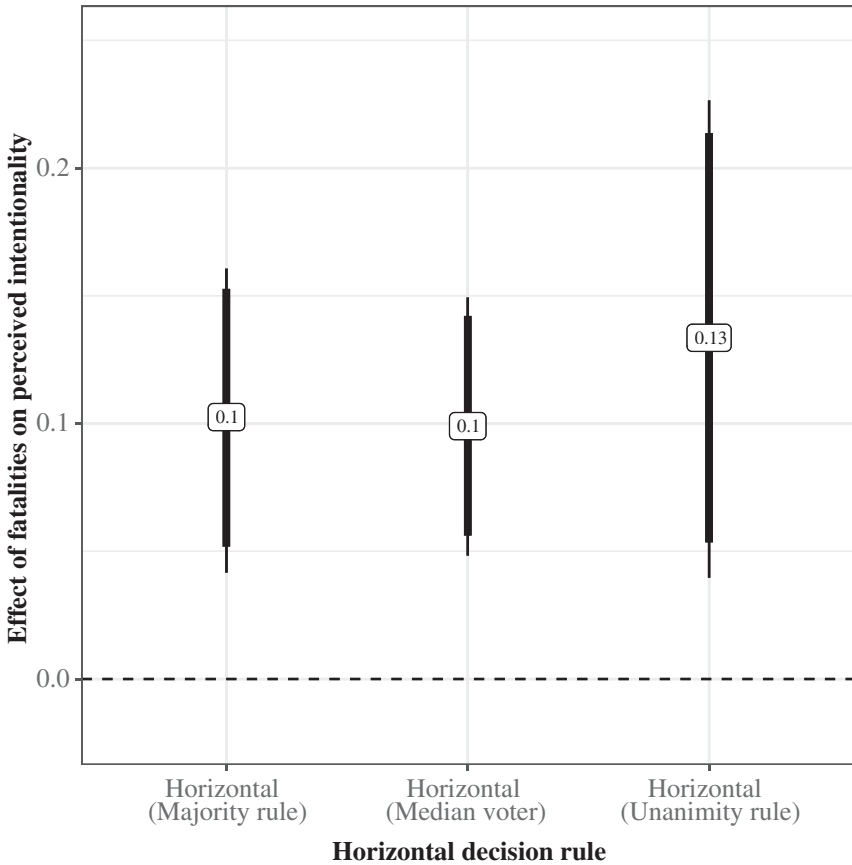


Notes: The figure shows the degree of perceived intentionality, given fatalities, within each group context (individual, hierarchical, or horizontal). The canonical intentionality bias result is seen in both hierarchical and horizontal groups. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals. Horizontal group decisions are calculated here using the median voter aggregation method. See Figure 6 for additional aggregation method results.

FIGURE 5. *Intentionality bias replicates in groups*

as a heuristic for value.⁵¹ Our treatment aims to test this second mechanism: American respondents might devalue a Chinese-authored proposal relative to an American-authored one because they would assume that the other country's negotiators do not have America's best interests in mind.

51. Brehm and Brehm 2013; Hovland and Weiss 1951; Ross 1993.

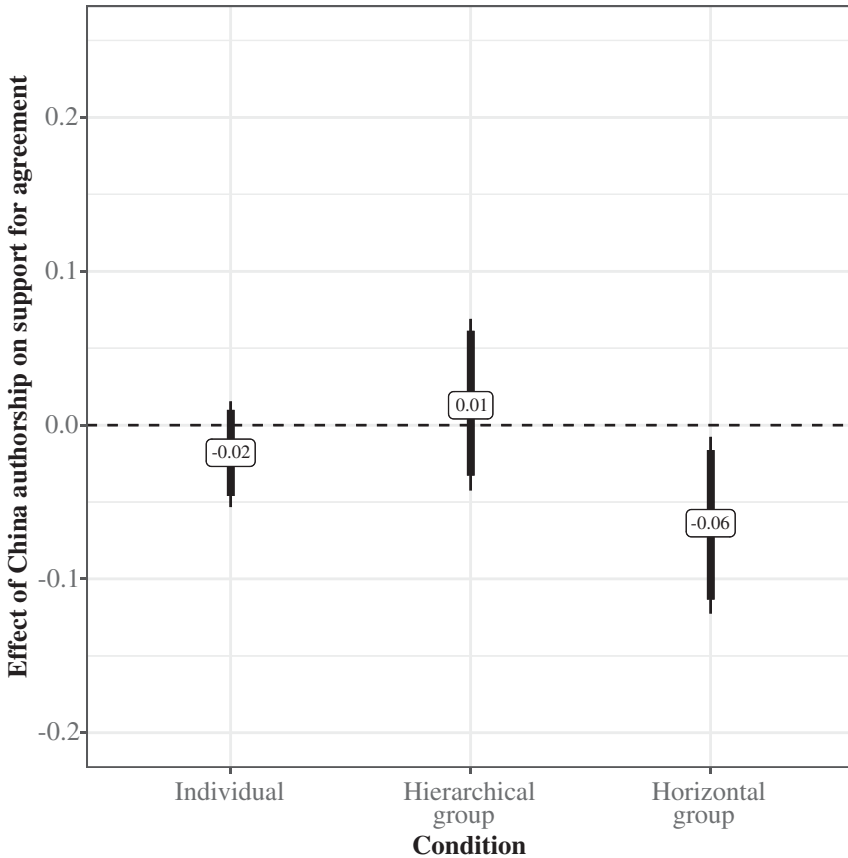


Notes: The figure displays the degree of perceived intentionality, given fatalities, within horizontal groups using different aggregation methods. The figure shows that the canonical intentionality bias result replicates across all three types of horizontal aggregation methods (majority rule, median voter, and unanimity rule). As in Figure 4, the tendency appears slightly larger in unanimous groups, but this difference is not statistically significant. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals.

FIGURE 6. *Intentionality bias effects by horizontal aggregation method*

However, to the extent that source credibility drives reactive devaluation, reactive devaluation should be strongest when individuals are presented with *ambiguous* proposals that increase their reliance on source heuristics.⁵² When the proposal is detailed and specific, subjects may be less likely to automatically devalue it because the proposal itself provides enough information to make an assessment. In

52. Maoz et al. 2002.



Notes: The figure shows the effect of Chinese (versus American) authorship on support for the policy proposal within each group context (individual, hierarchical, or horizontal). We fail to replicate the canonical reactive devaluation result in the individual and hierarchical group conditions, but do replicate it in the horizontal condition. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals. Horizontal group decisions are calculated here using the median voter aggregation method. See Figure 8 for additional aggregation method results.

FIGURE 7. *Reactive devaluation experiment displays mixed results*

our study, the proposal was quite specific and detailed, with bullet points outlining the exact compromises each side would make in the ongoing trade war. This level of detail may have attenuated reactive devaluation, making it easier for subjects to look past the purported authorship of the proposal to evaluate the actual proposal content.

Another possibility is that the conflict tested in this study—contested trade negotiations in the shadow of Trump-era trade wars—resulted in less reactive devaluation

either because of the unusual domestic politics of the Trump era, or simply because the rivalry was less clear-cut than the violent, intractable conflicts in which this bias has historically been studied. In other words, Israelis may be more suspicious and distrusting of Palestinians, and Americans more distrusting of the Soviet Union or North Vietnam during the Cold War, than Americans in 2020 were of China, with whom the United States had a less directly confrontational relationship.⁵³

However, even with the specificity of this proposal and ambiguity in the rivalry, we do observe reactive devaluation in horizontal groups, particularly those that reached unanimous decisions (Figure 8). Unanimous horizontal groups are marginally more likely than individuals ($p < .06$) to devalue the Chinese proposal relative to the American one. This suggests that, to the extent that the potential for reactive devaluation occurs in this context, groups are, if anything, increasing this tendency.

Extensions and Limitations

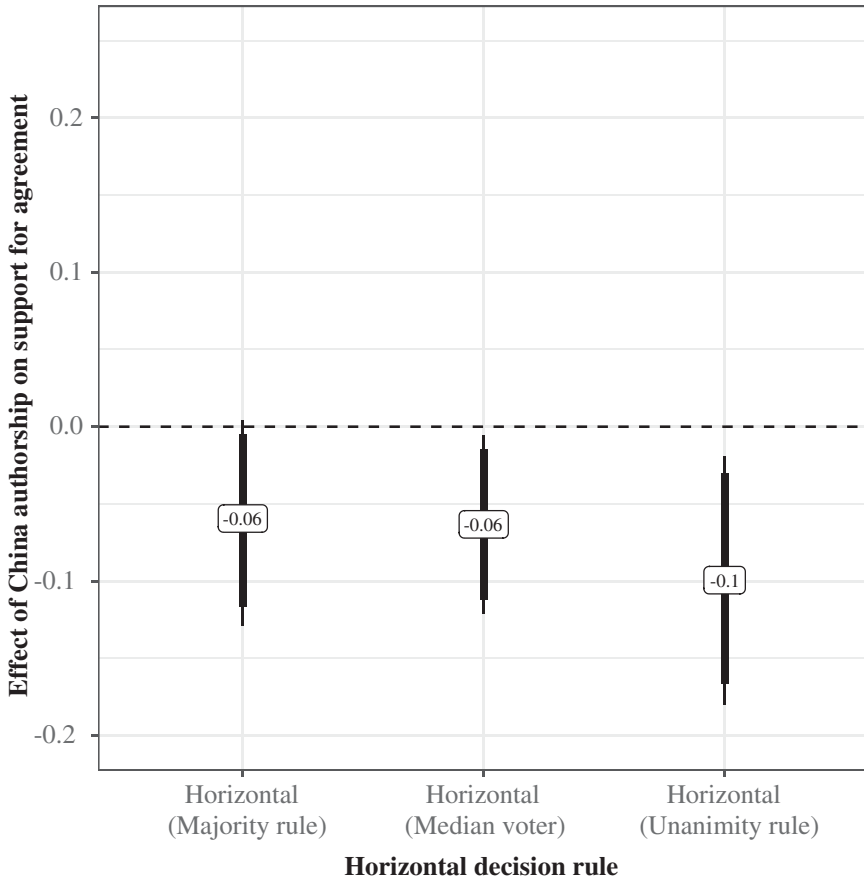
Thus far, our results suggest that two canonical biases from the judgment and decision-making literature—sensitivity to framing effects in prospect theory, and the intentionality bias—persist or become even more pronounced in group settings. And, while we fail to replicate reactive devaluation in our individual condition and hierarchical group contexts, we replicate it in horizontal groups, which is inconsistent with the claim that the hawkish biases that manifest in individual settings disappear in groups. However, there are important limitations and caveats worth discussing, many of which involve questions of external validity, and differences between inevitably stylized experiments and real-world foreign policy decision making.

First, our experiments lack many of the social dynamics of real foreign policy decision-making groups where there is social pressure, people have worked with each other before (and might again), issue linkage is possible, bureaucratic interests are present, and so on.⁵⁴ In contrast, our respondents participate anonymously, in novel groups formed explicitly for this study, with little social pressure for cohesion or prospect of future interaction.⁵⁵ We encourage future researchers to build on these studies by incorporating some of these features into their experimental designs to determine the impact of differing levels of social pressure on group susceptibility to bias. And yet the absence of these features likely makes our findings a more *conservative* test of groups' ability to reduce bias, since the features missing from our studies are also the very features typically linked to biased information-processing

53. Ashmore et al. 1979; Maoz et al. 2002; Ross and Ward 1995.

54. Allison 1971.

55. Although the fact that respondents complete multiple experimental modules in the same groups means that there is some opportunity for repeated interaction and social learning—and we do not find that the magnitude of the bias in our data decays over multiple experimental interactions—as a test of social pressure it is relatively modest.



Notes: The figure shows the effect of Chinese (versus American) authorship on support for the policy proposal within horizontal groups using different aggregation methods. The canonical reactive devaluation result is seen in two of the three aggregation methods (median voter and unanimity). As with the other two experiments, this tendency appears to be larger for unanimous groups. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals.

FIGURE 8. *Reactive devaluation effects by horizontal aggregation method*

and pathological group dynamics such as groupthink.⁵⁶ In that sense, the fact that we replicate the prospect theory and intentionality bias effects across all our group conditions even without the distorting effects of social conformity pressures should increase our confidence in the pervasiveness of these tendencies.

56. Janis 1972.

Second, in the real world, leaders are not randomly assigned but strategically selected for particular skills, attributes, or experience. On the one hand, this is precisely why experiments are helpful: in a naturalistic setting, it would be difficult to identify the effect of group structures independently of the properties of actors in specific roles in the group. Experiments, in contrast, let us harness the power of random assignment and sidestep these concerns about endogeneity. On the other hand, this also leads to an important empirical question: are groups with certain types of leaders better able to avoid these biases?

To test this question, we take advantage of the lengthy battery of individual differences administered to respondents at the beginning of the study. Since there are many potential traits that could moderate the impacts of framing effects, intentionality bias, and reactive devaluation, we adopt a data-driven approach, estimating a sparse Bayesian method for variable selection. We fit a LASSOplus model regressing our dependent variable on the treatment, a vector of twenty-one individual differences (foreign policy orientations, personality traits, demographic characteristics, government experience, and so on), and interactions between these leader-level traits and the treatment using data from the hierarchical condition.⁵⁷ This machine-learning approach thus lets us test whether certain kinds of leaders (such as those high in need for cognition, or with more experience) better help their groups avoid these biases. Crucially, none of these leader-level characteristics significantly moderate the treatments. We thus find no evidence that groups with better leaders are less likely to display these patterns. We encourage future work to build on these findings by assigning respondents with specific traits (such as narcissism) to leader and adviser roles, to test how it affects the quality of decision making.⁵⁸

The question of leader traits raises a related issue. Our study was conducted on samples of ordinary citizens, rather than experienced decision makers. It is of course possible that groups composed of actual elite decision makers would behave differently, though two considerations are relevant here. One is that these three hawkish biases have previously all been identified in foreign policy elites using archival and case study evidence,⁵⁹ so we already have reason to believe that foreign policy decision makers experience hawkish biases; the question is whether group contexts moderate the magnitude of these biases at a significantly different rate among elites than they do among members of the mass public. The other is that meta-analyses of paired experiments on elite and mass samples suggest strikingly similar responses to experimental treatments, so we should not assume that they rely on fundamentally different cognitive architectures.⁶⁰ Ultimately, however, this is an empirical question. It is also one that elite experiments may be poorly equipped to answer, suggesting benefits for archival or mixed-method approaches. Experimental or survey-based studies

57. Ratkovic and Tingley 2017.

58. Harden 2021.

59. McDermott 1998; Ross and Ward 1995; Traven 2021.

60. Kertzer 2021.

on real foreign policy decision makers invariably involve smaller sample sizes—effectively made smaller still once analyzed at the group level—such that many group-level elite experiments would likely be underpowered, particularly if they use the sample of elites most directly implicated by their theory.⁶¹

Group-Level Diversity

Yet even if leader-level traits don't seem to minimize these three biases, it is possible that group-level ones do. One of the most-studied attributes of groups hypothesized to improve decision making is diversity.⁶² Diversity refers most broadly to “compositional differences among people” within a particular unit, such as a decision-making group.⁶³ In a decision-making context, these compositional differences are often understood as representing the interaction of different cognitive styles. As Page has argued, in the context of problem solving for example, diversity of perspectives, interpretations, heuristics, and individual predictive models that are used to infer cause and effect all come together to “increase the number of solutions that a collection of people can find by creating different connections among the possible solutions.”⁶⁴ Diverse groups are also thought to lead to more extensive debate, increase exposure to others' viewpoints, introduce differences in risk preferences, and avoid group pathologies such as groupthink, where striving for uniformity may overwhelm accuracy motives.⁶⁵ In short, “diversity trumps homogeneity.”⁶⁶ Yet, groups that are too diverse may move too far in the other direction, to where a “polythink” dynamic prevents them from reaching consensus at all.⁶⁷ Relatedly, in some instances diverse groups may be more prone to conflict, as social identity and categorization processes may impede the value of information and perspective pooling that leads to higher group performance.⁶⁸

We therefore examine the potential mitigating effect of diversity on susceptibility to bias, assessing whether groups with a more diverse composition are affected less by these various hawkish biases. Rather than using Herfindahl indices, which flatten diversity onto a single dimension, we operationalize diversity in a multidimensional fashion, calculating the group-level variance of a given trait in each group, and averaging across diversity scores for four types of traits, to produce measures of four different types of diversity.

We first examine diversity from a demographic perspective, in which more diverse groups are those with members with different ages, gender and racial identities,

61. Kertzer and Renshon 2022.

62. Horowitz et al. 2019; Page 2007, 2019.

63. Roberson 2019, 70.

64. Page 2007, 9.

65. Janis 1972.

66. Page 2007, 10.

67. Mintz and Wayne 2016.

68. Roberson 2019.

religions, and socioeconomic backgrounds. This type of descriptive diversity, in addition to being normatively valuable, has been hypothesized to improve decision making by broadening the information set and policy options reviewed and considered by a group.⁶⁹ Second, we operationalize diversity in terms of personal dispositions: the “big five” personality characteristics, need for cognition, trait aggression, and risk orientation. This type of cognitive diversity is often studied in the organizational behavior literature, which is interested in how the variability of personality characteristics in teams affects their collective performance.⁷⁰ Third, we turn to diversity of experience within groups, where different members of the group have varying levels of experience in leadership and small-group decision making (political or otherwise). In foreign policy decision-making contexts, diversity of experience may be particularly important, since decision units are typically a mixture of experienced bureaucrats and shorter-term political figures, themselves with varying experience in government.⁷¹ Finally, we consider groups whose members vary in their political attitudes or orientations, including political ideology, right-wing authoritarianism, social dominance orientation, and foreign policy orientations. These types of attributes have long been theorized to play a prominent role in foreign policy beliefs and attitudes, but how the variance of these traits within a decision-making *unit* affects decision outcomes has been less explored.⁷²

Regardless of how we operationalize diversity, however, we find no systematic effects of diversity on susceptibility to any of the hawkish biases we examine. Diverse groups are just as likely as more homogeneous groups, and no less likely than individuals, to exhibit these biases (Figure 9).⁷³ It is not that diversity has no effects whatsoever: more diverse groups, particularly those with more diverse dispositions and political attitudes, are more likely to fail to reach agreement at all (Figure 10). This is particularly the case in the intentionality bias and reactive devaluation experiments, where respondents are assessing adversarial interactions with China and North Korea. Groups whose members hold different social and political attitudes are more likely to show internal dissensus and disagreement.⁷⁴ Nonetheless, more diverse groups do not appear to be less likely to display these three tendencies.

69. Page 2019.

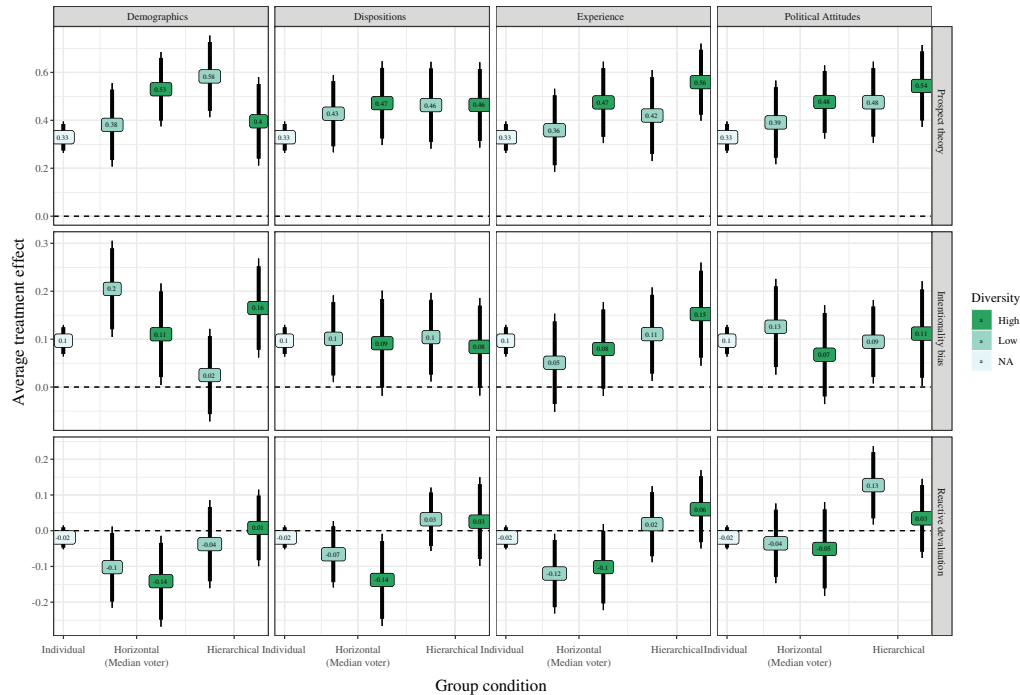
70. Halfhill et al. 2005.

71. Mintz, Redd, and Vedlitz 2006; Saunders 2017.

72. Hermann 2001; Larson 1994; Rathbun et al. 2016.

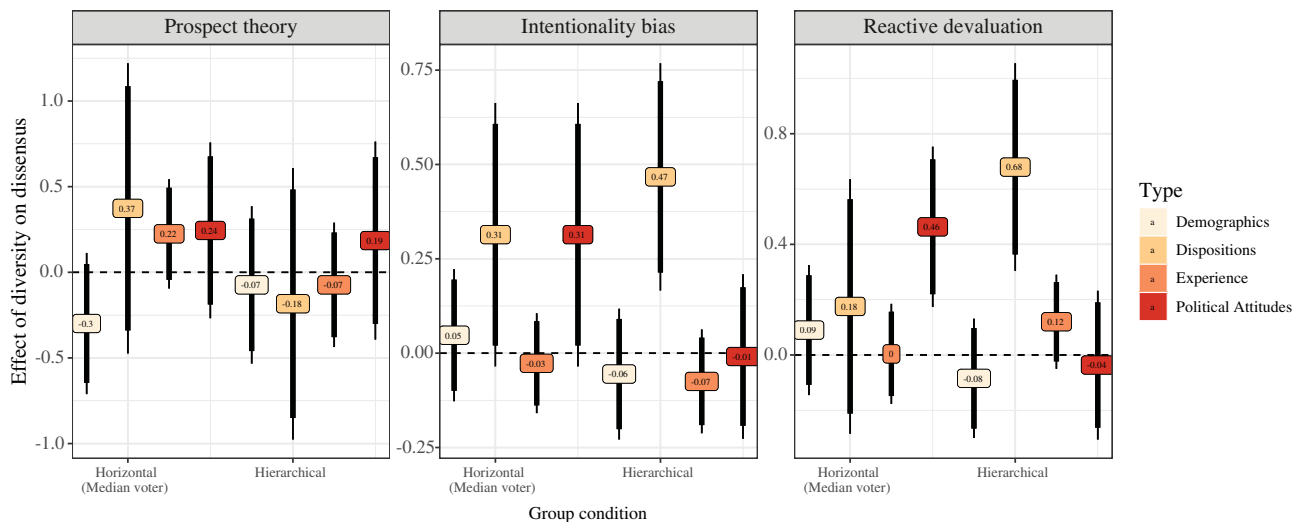
73. As a robustness check, we also examine the effect of gender composition in groups in particular, in both absolute terms (e.g., the *number* of group members who do not identify as male) and relative ones (the *proportion* who do not identify as male). The number of non-male group members in the horizontal condition appears to moderate the treatment regardless of the functional form used, and the LASSOplus results suggest that leader gender does not meaningfully affect group decisions in the hierarchical condition either.

74. The dissensus measure is the variance in the dependent variable among members of the group. Greater variance among group members in the preferred decision is more dissensus.



Notes: The figure explores the effects of group diversity on susceptibility to hawkish biases by comparing the average treatment effect for less diverse (25th percentile and below) and more diverse (75th percentile and above) groups in the horizontal and hierarchical conditions, benchmarking each result with the average treatment from the individual condition. Each row depicts the result from a different experimental module, and each column operationalizes diversity using a different metric (based on group members' demographics, dispositions, prior experiences, or political attitudes). We find that (1) more diverse groups are not significantly less prone to hawkish biases than less diverse groups (borne out by more formal interaction models); and (2) more diverse groups are not significantly less prone than individuals to hawkish biases. Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals. Horizontal group decisions are calculated here using the median voter aggregation method.

FIGURE 9. *More diverse groups are no less susceptible to these three biases*



Notes: The figure considers the effects of group diversity on dissensus within each group condition. Each panel depicts the result from a different experimental module, operationalizing diversity using our four different metrics (based on group members' demographics, dispositions, experiences, or political attitudes). More diverse groups seem more likely to experience dissensus (especially in the intentionality bias and reactive devaluation experiments). Point estimates are cell means with 90 and 95 percent bootstrapped confidence intervals. Horizontal group decisions are calculated here using the median voter decision rule.

FIGURE 10. *More diverse groups are more likely to experience dissension*

Group Size and Modes of Interaction

Finally, there are two other considerations worth noting, which also serve as alternative interpretations of our results. One is that for the ameliorative effects of aggregation to take place, group members need to interact face to face rather than deliberate at a distance.⁷⁵ Another is that for the ameliorative effects of aggregation to take place, groups need to be much larger; after all, foreign ministries have hundreds or thousands of individuals. While small groups might replicate individual-level biases, the “wisdom of crowds” might suggest greater rationality as groups grow in size.⁷⁶ On the one hand, these interpretations are obviously in tension with one another, since as groups increase in size, the rate of face-to-face communication decreases. On the other, there are a number of empirical tests we can employ to speak to some of these questions directly.

First, we can exploit variation in group size in our results. The magnitude of the hawkish biases we observe does not significantly shrink with group size (see Section 3 in the online supplement), and simulation methods suggest that some might actually increase.

This pattern comports with archival evidence from the United States regarding leaders’ frustrations with the pathologies of large decision-making units and the perception that larger groups had more problematic tendencies than smaller ones. As a result, while there was variation from administration to administration, a number of high-profile decisions, from the Cuban Missile Crisis to the first Gulf War, often involve the president and a relatively small number of influential advisers.⁷⁷ John F. Kennedy, for example, was disappointed by the results of relying on a large number of advisers, noting, “The advice of every member of the Executive Branch brought in to advise was unanimous—and the advice was wrong.” In response, at least partially, to these perceived failings of larger groups, Kennedy created a smaller Executive Committee, and often relied on ad hoc meetings of even smaller groups within it. Similarly, George H.W. Bush relied on ad hoc small groups of advisers when deciding whether to invade Iraq. The results from this study are likely directly applicable to these types of cases of relatively small group decision making, which have been quite common in historical US foreign policymaking.

Second, although all our respondents participated online rather than in person, if we think about face-to-face interaction in terms of the added information it conveys, we can test this informational mechanism directly by testing whether groups where respondents exchanged more information as part of their deliberations displayed weaker biases than groups where respondents communicated less.⁷⁸ Interestingly, across all three experiments, for both horizontal and hierarchical

75. Holmes 2018.

76. Surowiecki 2005.

77. For a review, see Jordan et al. 2009.

78. Holmes 2018.

groups, we find no evidence that the magnitude of the biases groups display significantly decreases with group participation (see section 4 in the online supplement).⁷⁹

One explanation may relate to behavioral modifications that are made when more information-rich environments, such as face-to-face meetings, are unavailable. When unable to communicate with visual expressive behaviors, individuals use textual proxies for visual cues, which in some cases may enhance, rather than degrade, social bonding processes.⁸⁰ Research in social information processing theory suggests that when individuals meet for the first time, as is the case in our study, text-based communication can enhance intimacy and self-disclosure, positively affecting relationship building.⁸¹ For example, Wheeler and Holmes argue that face-to-face interaction as a quotidian practice of international politics is a relatively recent phenomenon, which means that text-based communication was, historically, the only route to relationship building.⁸² Particularly as global pandemics take diplomacy online, we see questions about the role of interaction modality in group decision-making as an important topic for future research.

Conclusion

In a recent review of the problem of aggregation, Gildea notes that “how psychological mechanisms, which are primarily individually embodied, may operate and exercise influence within complex group and institutional environments remains a crucial and contested question.”⁸³ To date, such concerns have remained largely conceptual in nature, and the answer to this question has proven elusive because studying it empirically introduces a number of difficult methodological and substantive challenges. We offer a direct test of how a particular class of psychological biases aggregate in foreign policy contexts by experimentally testing how a trio of so-called “hawkish biases” linked to foreign policy aggregate in groups. Our results, which suggest that the aggregation problem may be less problematic than some scholars have alleged, and that individual-level psychological biases do not necessarily cancel out in groups, may be surprising for some. If “the whole point of government is to ensure multiple voices and checks and balances so that rational decisions can, in theory, persist despite individual preferences and biases,” we may need to revisit the assumption that multiple voices lead to more rational outcomes.⁸⁴ Our results suggest that the biases that manifest in lone voices are similarly present in group decision making.

79. Importantly, these tests also suggest that our replication of these biases in the group conditions is unlikely to be an artifact of group members’ not taking the study seriously.

80. Walther 1992.

81. Antheunis, Valkenburg, and Peter 2007; Tidwell and Walther 2006.

82. Wheeler and Holmes 2021.

83. Gildea 2020, 1–2.

84. D. Johnson 2015, 760.

One important theoretical implication of our findings is that we should be more comfortable envisioning individual-level biases scaling up to small groups in decision-making contexts. In an important application of prospect theory to foreign policy, McDermott applied the bias to a number of cases, focusing “on a unitary actor embodied by the president.” She notes that “prospect theory is less easily applied to the dynamics of *group* decision making, except to the extent that all members are assumed to share similar biases in risk propensity, although each may possess a different understanding of such crucial features as appropriate frame for discussion, applicable reference point, domain of action, and so on.”⁸⁵ By analyzing prospect theory’s applicability to groups experimentally, we are able to control many of these elements, including the domain of action and parameters for discussion, and our results suggest that such an application of individual psychology to groups may therefore not be as infeasible as some may fear. Further empirical work is required to assess how the experimental results we obtain here generalize to those in historical cases, while additional experimental work will likely be helpful in establishing how the group decision-making process operates. One such question concerns the study of reference point in groups. As Kameda and Davis ask, “What happens if a group is composed of some members who have experienced certain losses recently and others who have experienced certain gains recently?”⁸⁶ Randomly assigning group members with treatments that condition their individual reference points may allow researchers to trace the effects of those reference points in the group decision-making process.

An additional potential implication concerns our failure to detect beneficial effects of diversity on group decision making. One reason for this may relate to the nature of the tasks we employ here: unlike the protocols used in many of the experimental tests in the wisdom-of-the-crowd literature, testing the “miracle of aggregation” using math problems or prediction tasks, none of these studies have an objective right answer. In this sense, though, they better resemble the ill-structured problems that characterize much of foreign policy decision making, suggesting that the wisdom of the crowd may be a poor analogy for many of the questions IR scholars care about—although we also examine this question directly in follow-up work, using incentivized group bargaining experiments.⁸⁷ Future research should also focus on identifying other possible diversity mechanisms, such as those that relate to visible diversity and face-to-face interactions.⁸⁸ In face-to-face contexts, group members will likely be more aware of diversity within their group, creating a possibility that group members’ knowledge of their group’s diversity affects their problem solving.

Another interpretation may have to do with the robustness of the biases themselves. Perhaps the three cognitive biases examined in this study are particularly ubiquitous

85. McDermott 1998, 187.

86. Kameda and Davis 1990, 58.

87. Brutger and Kertzer 2018.

88. E.g., Staples and Zhao 2006.

and resistant to attempts at mitigation. We have some empirical evidence on this front: we use the same LASSOplus approach we used in the leader characteristic analysis, but testing for heterogeneous treatment effects by individual-level traits in the individual condition. As before, none of these individual differences significantly moderate the treatments. Thus, one potential reason why we fail to find that diversity has mitigating effects has to do with the robustness of the regularities we study here. In other words, diversity may be beneficial in improving decision making in other crucial ways, even if it does not appreciably alter a group's susceptibility to these types of cognitive biases.⁸⁹ Yet the fact that these "nonstandard preferences" appear to be so robust also suggests the merits of rational choice approaches incorporating these regularities into their models.⁹⁰ In other experimental work, we build on these findings by examining how individual-specific traits relevant to foreign policy decision making—rather than these judgment and decision-making biases that appear to be fairly robust across individuals—aggregate in group decision-making contexts.

This is not to say that groups do not exhibit their own peculiarities that may lead to subrational or irrational outcomes. It may be, for example, that not only do groups not reduce the effects of cognitive biases, they introduce new dynamics that may exacerbate deviations from expected utility models. Early psychological research identified many of these tendencies. "Risky shifts," or the tendency of individuals in groups to make riskier decisions than when polled individually, is a finding that led to a robust literature on group polarization, consistent with the findings of our prospect theory experiment.⁹¹ Similarly, initial studies on group conformity spurred over half a century of investigating the conditions in which groups create conformity dynamics in foreign policy situations, particularly as they relate to perceived policy failures.⁹² It may be, however, that groupthink is receiving unfair blame. As Whyte has argued, "history and the daily newspaper provide examples of policy decisions made by groups that resulted in fiascos. The making of such decisions is frequently attributed to the groupthink phenomenon"—though it may be that "prospect polarization" instead is the culprit.⁹³ Precisely because cognitive biases have largely been studied at the individual level, and not believed to be a group-level phenomenon, group-level theories such as groupthink have taken on a heavy explanatory burden. By relaxing the assumption that we need group-level theories to explain "nonstandard decision making," new explanatory frameworks become available. It is also conceivable that the persistence of cognitive biases in groups exacerbates conformity dynamics by facilitating premature consensus, a possibility worthy of future research.

89. And, of course, descriptive diversity can be normatively valuable regardless of any benefits it may provide for decision making.

90. Kertzer 2016; Mintz, Valentino, and Wayne 2021; Stein 2017.

91. Stoner 1961.

92. Asch 1951; Badie 2010; Janis 1972; Sherif 1935.

93. Whyte 1989, 40.

Finally, while our focus here is on the aggregation of biases that IR scholars have argued are particularly important in foreign policy decision making, it is worth noting that our findings are relevant for the study of collective decision making in a wide range of contexts. Prospect theory is frequently applied to a variety of questions in American and comparative politics; intentionality bias is central to questions of blame attribution in politics more generally; and reactive devaluation is tightly linked to theories of negative partisanship.⁹⁴ These findings should therefore be of interest to scholars of collective decision making across a broad set of domestic political issues, rather than just foreign ones.

In treating aggregation as an empirical rather than conceptual question, our study also has important implications beyond the three biases studied here. While we focused on studying group decision making in the context of foreign policy, similar group processes are present in a wide range of complex institutional environments. Practice theorists, for example, have argued that diplomacy in an organization such as NATO includes micro dyadic interactions between individual diplomats, as well as collective decision making in which diplomats conform with logics of practice or habit.⁹⁵ During NATO decision-making sessions on the proposed use of force in Libya in 2011, for example, Adler-Nissen and Pouliot report that diplomats drew on the taken-for-granted nature of the decision making, noting that “at some point you just know where the wind blows,” and that in these discussions, “the diplomatic process gradually gains a life of its own.”⁹⁶ One of the criticisms levied at this type of approach, however, is that the mechanism by which a group comes to know which way the wind is blowing, or how diplomacy gains a life of its own, is often underspecified, making it difficult to know a priori when and what types of practices are likely to affect outcomes in any given setting.⁹⁷ Our methodological approach offers one step toward a potential solution. By studying aggregation empirically, group experiments such as those reported here may help us better identify the ways in which group practical sense is created, providing an incremental step in building microfoundations for practice theories. Altogether, this research shows the value of treating the “aggregation problem” in foreign policy as a phenomenon that deserves to be studied empirically, rather than just assumed.

Data Availability Statement

Replication files for this article may be found at <<https://doi.org/10.7910/DVN/N8GBLF>>.

94. E.g., Brutger 2021; Malhotra and Kuo 2008; McDermott 2004; McGraw 1991; Sheffer et al. 2018.

95. McCourt 2016.

96. Adler-Nissen and Pouliot 2014.

97. Ringmar 2014, 6.

Supplementary Material

Supplementary material for this article is available at <https://doi.org/10.1017/S0020818322000017>.

References

- Adler-Nissen, Rebecca, and Vincent Pouliot. 2014. Power in Practice: Negotiating the International Intervention in Libya. *European Journal of International Relations* 20 (4):889–911.
- Allison, Graham. 1971. *Essence of Decision*. Little, Brown and Company.
- Antheunis, Marjolijn L., Patti M. Valkenburg, and Jochen Peter. 2007. Computer-Mediated Communication and Interpersonal Attraction: An Experimental Test of Two Explanatory Hypotheses. *Cyberpsychology and Behavior* 10 (6):831–36.
- Arrow, Kenneth J. 1950. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* 58 (4):328–46.
- Asch, Solomon E. 1951. Effects of Group Pressure upon the Modification and Distortion of Judgments. In *Groups, Leadership and Men: Research in Human Relations*, edited by Harold Guetzkow, 177–90. Carnegie Press
- Ashmore, Richard D., David Bird, Frances K. Del Boca, and Robert C. Vanderet. 1979. An Experimental Investigation of the Double Standard in the Perception of International Affairs. *Political Behavior* 1 (2): 123–35.
- Ausderan, Jacob Thomas. 2013. *International Conflict and the Strategic Selection of Foreign Policy Advisors*. PhD diss., Florida State University.
- Austen-Smith, David, and Jeffrey S. Banks. 1996. Information Aggregation, Rationality, and the Condorcet Jury Theorem. *American Political Science Review* 90 (1):34–45.
- Badie, Dina. 2010. Groupthink, Iraq, and the War on Terror: Explaining US Policy Shift Toward Iraq. *Foreign Policy Analysis* 6 (4):277–96.
- Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen. 2019. The Role of Evidence in Politics: Motivated Reasoning and Persuasion Among Politicians. *British Journal of Political Science* 49 (3):1117–40.
- Berdahl, Andrew, Colin J. Torney, Christos C. Ioannou, Jolyon J. Faria, and Iain D. Couzin. 2013. Emergent Sensing of Complex Environments by Mobile Animal Groups. *Science* 339 (6119):574–76.
- Bernstein, Ethan, Jesse Shore, and David Lazer. 2018. How Intermittent Breaks in Interaction Improve Collective Intelligence. *Proceedings of the National Academy of Sciences* 115 (35):8734–39.
- Brehm, Sharon S., and Jack W. Brehm. 2013. *Psychological Reactance: A Theory of Freedom and Control*. Academic Press.
- Brooks, Sarah M., Raphael Cunha, and Layna Mosley. 2015. Categories, Creditworthiness, and Contagion: How Investors' Shortcuts Affect Sovereign Debt Markets. *International Studies Quarterly* 59 (3):587–601.
- Brutger, Ryan. 2021. The Power of Compromise: Proposal Power, Partisanship, and Public Support in International Bargaining. *World Politics* 73 (1):128–66.
- Brutger, Ryan, and Joshua D. Kertzer. 2018. A Dispositional Theory of Reputation Costs. *International Organization* 72 (3):693–724.
- Chu, Jonathan, Marcus Holmes, and David Traven. 2021. Intentions from Consequences: How Moral Judgments Shape Citizen Perceptions of Wartime Conduct. *Journal of Experimental Political Science* 8 (2):203–207.
- Davis, James W., and Rose McDermott. 2021. The Past, Present, and Future of Behavioral IR. *International Organization* 75 (1):147–77.
- DellaVigna, Stefano. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* 47 (2):315–72.

- Devine, Dennis J., Laura D. Clayton, Benjamin B. Dunford, Rasmy Seyer, and Jennifer Pryce. 2001. Jury Decision Making: Forty-five Years of Empirical Research on Deliberating Groups. *Psychology, Public Policy, and Law* 7 (3):622–727.
- Esser, James K. 1998. Alive and Well After Twenty-five Years: A Review of Groupthink Research. *Organizational Behavior and Human Decision Processes* 73 (2–3):116–41.
- Etheredge, Lloyd S. 1978. Personality Effects on American Foreign Policy, 1898–1968: A Test of Interpersonal Generalization Theory. *American Political Science Review* 72 (2):434–51.
- George, Alexander L. 1969. The “Operational Code”: A Neglected Approach to the Study of Political Leaders and Decision Making. *International Studies Quarterly* 13 (2):190–222.
- George, Alexander L. 1972. The Case for Multiple Advocacy in Making Foreign Policy. *American Political Science Review* 66 (3):751–85.
- Gigerenzer, Gerd, and Wolfgang Gaissmaier. 2011. Heuristic Decision Making. *Annual Review of Psychology* 62:451–82.
- Gildea, Ross James. 2020. Psychology and Aggregation in International Relations. *European Journal of International Relations* 26 (S1):166–83.
- Greenstein, Fred I. 1969. *Personality and Politics: Problems of Evidence, Inference, and Conceptualization*. Markham.
- Grynaviski, Eric. 2014. *Constructive Illusions: Misperceiving the Origins of International Cooperation*. Cornell University Press.
- Hafner-Burton, Emilie M., Stephan Haggard, David A. Lake, and David G. Victor. 2017. The Behavioral Revolution and International Relations. *International Organization* 71 (S1):S1–S31.
- Halfhill, Terry, Eric Sundstrom, Jessica Lahner, Wilma Calderone, and Tjai M. Nielsen. 2005. Group Personality Composition and Group Effectiveness: An Integrative Review of Empirical Research. *Small Group Research* 36 (1):83–105.
- Harden, John P. 2021. All the World’s a Stage: US Presidential Narcissism and International Conflict. *International Studies Quarterly* 65 (3):825–37.
- Hart, Paul ‘t, Eric Stern, and Bengt Sundelius. 1997. *Beyond Groupthink: Political Group Dynamics and Foreign Policy-Making*. University of Michigan Press.
- Hermann, Margaret G. 1980. Explaining Foreign Policy Behavior Using the Personal Characteristics of Political Leaders. *International Studies Quarterly* 24 (1):7–46.
- Hermann, Margaret G. 2001. How Decision Units Shape Foreign Policy: A Theoretical Framework. *International Studies Review* 3 (2):47–81.
- Hermann, Margaret G., and Thomas Preston. 1994. Presidents, Advisers, and Foreign Policy: The Effect of Leadership Style on Executive Arrangements. *Political Psychology* 15 (1):75–96.
- Holmes, Marcus. 2018. *Face-to-Face Diplomacy: Social Neuroscience and International Relations*. Cambridge University Press.
- Horowitz, Michael, Brandon M. Stewart, Dustin Tingley, Michael Bishop, Laura Resnick Samotin, Margaret Roberts, Welton Chang, Barbara Mellers, and Philip Tetlock. 2019. What Makes Foreign Policy Teams Tick: Explaining Variation in Group Performance at Geopolitical Forecasting. *Journal of Politics* 81 (4):1388–404.
- Horowitz, Michael C., and Matthew Fuhrmann. 2018. Studying Leaders and Military Conflict: Conceptual Framework and Research Agenda. *Journal of Conflict Resolution* 62 (10):2072–86.
- Hovland, Carl I., and Walter Weiss. 1951. The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly* 15 (4):635–50.
- Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Houghton Mifflin.
- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.
- Johnson, Dominic D.P. 2015. Survival of the Disciplines: Is International Relations Fit for the New Millennium? *Millennium* 43 (2):749–63.
- Johnson, Dominic D.P. 2020. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics*. Princeton University Press.
- Johnson, Richard Tanner. 1974. *Managing the White House: An Intimate study of the Presidency*. Harper Collins.

- Jordan, Amos A., William J. Taylor Jr., Michael J. Meese, and Suzanne C. Nielsen. 2009. *American National Security*. Johns Hopkins University Press.
- Kaarbo, Juliet. 1998. Power Politics in Foreign Policy: The Influence of Bureaucratic Minorities. *European Journal of International Relations* 4 (1):67–97.
- Kahneman, Daniel, and Jonathan Renshon. 2007. Why Hawks Win. *Foreign Policy* 158:34–38.
- Kahneman, Daniel, and Jonathan Renshon. 2009. Hawkish Biases. In *American Foreign Policy and the Politics of Fear: Threat Inflation Since 9/11*, edited by A. Trevor Thrall and Jane K. Cramer, 79–96. Routledge.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47 (2):263–91.
- Kahneman, Daniel, and Amos Tversky. 2017. Conflict Resolution: A Cognitive Perspective. In *Preference, Belief, and Similarity: Selected Writings of Amos Tversky*, edited by Eldar Shafir, 729–46. MIT Press.
- Kameda, Tatsuya, and James H. Davis. 1990. The Function of the Reference Point in Individual and Group Risk Decision Making. *Organizational Behavior and Human Decision Processes* 46 (1):55–76.
- Kerr, Norbert L., Robert J. MacCoun, and Geoffrey P. Kramer. 1996. Bias in Judgment: Comparing Individuals and Groups. *Psychological Review* 103 (4):687–719.
- Kerr, Norbert, and Scott Tindale. 2004. Group Performance and Decision Making. *Annual Review of Psychology* 55:623–55.
- Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton University Press.
- Kertzer, Joshua D. 2021. Re-assessing Elite Public Gaps in Political Behavior. *American Journal of Political Science* (online first).
- Kertzer, Joshua D., Brian C. Rathbun, and Nina Srinivasan Rathbun. 2020. The Price of Peace: Motivated Reasoning and Costly Signaling in International Relations. *International Organization* 74 (1):95–118.
- Kertzer, Joshua D., and Jonathan Renshon. 2022. Experiments and Surveys on Political Elites. *Annual Review of Political Science* 25:1–26.
- Kertzer, Joshua D., and Dustin Tingley. 2018. Political Psychology in International Relations: Beyond the Paradigms. *Annual Review of Political Science* 21:319–39.
- Knobe, Joshua. 2003. Intentional Action and Side Effects in Ordinary Language. *Analysis* 63 (3):190–94.
- Lake, David A., and Robert Powell, eds. 1999. *Strategic Choice and International Relations*. Princeton University Press.
- Landau-Wells, Marika. 2018. Dealing with Danger: Threat Perception and Policy Preferences. PhD diss., Massachusetts Institute of Technology.
- Larrick, Richard P. 2016. The Social Context of Decisions. *Annual Review of Organizational Psychology and Organizational Behavior* 3 (1):441–67.
- Larson, Deborah Welch. 1985. *Origins of Containment: A Psychological Explanation*. Princeton University Press.
- Larson, Deborah Welch. 1994. The Role of Belief Systems and Schemas in Foreign Policy Decision-Making. *Political Psychology* 15 (1):17–33.
- Leites, Nathan. 1951. *The Operational Code of the Politburo*. McGraw-Hill.
- LeVeck, Brad L., D. Alex Hughes, James H. Fowler, Emilie Hafner-Burton, and David G. Victor. 2014. The Role of Self-Interest in Elite Bargaining. *Proceedings of the National Academy of Sciences* 111 (52):18536–41.
- LeVeck, Brad L., and Neil Narang. 2017. The Democratic Peace and the Wisdom of Crowds. *International Studies Quarterly* 61 (4):867–80.
- Levy, Jack S. 1996. Loss Aversion, Framing, and Bargaining: The Implications of Prospect Theory for International Conflict. *International Political Science Review* 17 (2):179–95.
- Levy, Jack S. 1997. Prospect Theory, Rational Choice, and International Relations. *International Studies Quarterly* 41 (1):87–112.
- Levy, Jack S. 2013. Psychology and Foreign Policy Decision-Making. In *Oxford Handbook of Political Psychology*, 2nd ed., edited by Leonie Huddy, David O. Sears, and Jack S. Levy, 301–33. Oxford University Press.

- Lewin, Kurt, Ronald Lippitt, and Ralph K. White. 1939. Patterns of Aggressive Behavior in Experimentally Created "Social Climates." *Journal of Social Psychology* 10 (2):269–99.
- Malhotra, Neil, and Alexander G. Kuo. 2008. Attributing Blame: The Public's Response to Hurricane Katrina. *Journal of Politics* 70 (1):120–35.
- Maoz, Ifat, Andrew Ward, Michael Katz, and Lee Ross. 2002. Reactive Devaluation of an "Israeli" vs. "Palestinian" Peace Proposal. *Journal of Conflict Resolution* 46 (4):515–46.
- Mattes, Michaela, and Jessica L.P. Weeks. 2019. Hawks, Doves and Peace: An Experimental Approach. *American Journal of Political Science* 63 (1):53–66.
- McCourt, David M. 2016. Practice Theory and Relationalism as the New Constructivism. *International Studies Quarterly* 60 (3):475–85.
- McDermott, Rose. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. University of Michigan Press.
- McDermott, Rose. 2004. Prospect Theory in Political Science: Gains and Losses from the First Decade. *Political Psychology* 25 (2):289–312.
- McGraw, Kathleen M. 1991. Managing Blame: An Experimental Test of the Effects of Political Accounts. *American Political Science Review* 85 (4):1133–57.
- Mercer, Jonathan. 1995. Anarchy and Identity. *International Organization* 49 (2):229–52.
- Mintz, Alex. 2007. Why Behavioral IR? *International Studies Review* 9 (1):157–62.
- Mintz, Alex, Steven B. Redd, and Arnold Vedlitz. 2006. Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations? *Journal of Conflict Resolution* 50 (5):757–76.
- Mintz, Alex, Nicholas Valentino, and Carly Wayne. 2021. *Beyond Rationality: Behavioral Political Science in the Twenty-first Century*. Cambridge University Press.
- Mintz, Alex, and Carly Wayne. 2016. *The Polythink Syndrome: US Foreign Policy Decisions on 9/11, Afghanistan, Iraq, Iran, Syria, and ISIS*. Stanford University Press.
- Mitzen, Jenniger, and Randall L. Schweller. 2011. Knowing the Unknown Unknowns: Misplaced Certainty and the Onset of War. *Security Studies* 20 (1):2–35.
- Molnar, Andras. 2019. SMARTRIQS: A Simple Method Allowing Real-Time Respondent Interaction in Qualtrics Surveys. *Journal of Behavioral and Experimental Finance* 22:161–69.
- Moynihan, Lisa M., and Randall S. Peterson. 2001. A Contingent Configuration Approach to Understanding the Role of Personality in Organizational Groups. *Research in Organizational Behavior* 23:327–78.
- Myers, David G., and Helmut Lamm. 1976. The Group Polarization Phenomenon. *Psychological Bulletin* 83 (4):602–27.
- Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- Page, Scott E. 2019. *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy*. Princeton University Press.
- Pechenkina, Anna O., and Nichole Argo. 2020. How Do Civilians Assign Blame and Praise Amidst Civil Conflict? *Behavioral Sciences of Terrorism and Political Aggression* 12 (4):243–67.
- Poulsen, Lauge N. Skovgaard, and Emma Aisbett. 2013. When the Claim Hits: Bilateral Investment Treaties and Bounded Rational Learning. *World Politics* 65 (2):273–313.
- Powell, Robert. 2017. Research Bets and Behavioral IR. *International Organization* 71 (S1):S265–S277.
- Powers, Kathleen E. 2022. *Nationalisms in International Politics*. Princeton University Press.
- Preston, Thomas. 2001. *The President and His Inner Circle: Leadership Style and the Advisory Process in Foreign Policy Making*. Columbia University Press.
- Rathbun, Brian C. 2014. *Diplomacy's Value: Creating Security in 1920s Europe and the Contemporary Middle East*. Cornell University Press.
- Rathbun, Brian C., Joshua D. Kertzer, Jason Reifler, Paul Goren, and Thomas J. Scotto. 2016. Taking Foreign Policy Personally: Personal Values and Foreign Policy Attitudes. *International Studies Quarterly* 60 (1):124–37.

- Ratkovic, Marc, and Dustin Tingley. 2017. Sparse Estimation and Uncertainty with Application to Subgroup Analysis. *Political Analysis* 25 (1):1–40.
- Redd, Steven B. 2002. The Influence of Advisers on Foreign Policy Decision Making: An Experimental Study. *Journal of Conflict Resolution* 46 (3):335–64.
- Renshon, Jonathan. 2017. *Fighting for Status: Hierarchy and Conflict in World Politics*. Princeton University Press.
- Ringmar, Erik. 2014. The Search for Dialogue as a Hindrance to Understanding: Practices as Inter-paradigmatic Research Program. *International Theory* 6 (1):1–27.
- Roberson, Quinetta M. 2019. Diversity in the Workplace: A Review, Synthesis, and Future Research Agenda. *Annual Review of Organizational Psychology and Organizational Behavior* 6 (1):69–88.
- Ross, Lee. 1993. *Reactive Devaluation in Negotiation and Conflict Resolution*. Stanford Center on International Conflict and Negotiation, Stanford University.
- Ross, Lee, and Andrew Ward. 1995. Psychological Barriers to Dispute Resolution. In *Advances in Experimental Social Psychology*, vol. 27, edited by M.P. Zanna, 255–304. Elsevier.
- Saunders, Elizabeth N. 2011. *Leaders at War: How Presidents Shape Military Interventions*. Cornell University Press.
- Saunders, Elizabeth N. 2017. No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making. *International Organization* 71 (S1):S219–S247.
- Schafer, Mark, and Scott Crichlow. 2010. *Groupthink Versus High-Quality Decision Making in International Relations*. Columbia University Press.
- Schneeweiss, Christoph. 2012. *Distributed Decision Making*. Springer Science and Business Media.
- Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave, and Tamir Sheafer. 2018. Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians. *American Political Science Review* 112 (2):302–21.
- Sherif, Muzafer. 1935. A Study of Some Social Factors in Perception. *Archives of Psychology*: 5–60.
- Staples, D., and Lina Zhao. 2006. The Effects of Cultural Diversity in Virtual Teams Versus Face-to-Face Teams. *Group Decision and Negotiation* 15 (July):389–406.
- Stein, Janice Gross. 1988. Building Politics into Psychology: The Misperception of Threat. *Political Psychology* 9 (2):245–71.
- Stein, Janice Gross. 2017. The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics. *International Organization* 71 (S1):S249–S263.
- Stoner, James Arthur Finch. 1961. A Comparison of Individual and Group Decisions Involving Risk. Master's thesis, Massachusetts Institute of Technology.
- Strodbeck, Fred L., Rita M. James, and Charles Hawkins. 1957. Social Status in Jury Deliberations. *American Sociological Review* 22 (6):713–19.
- Sunstein, Cass R., and Reid Hastie. 2014. Making Dumb Groups Smarter. *Harvard Business Review* 92 (12):90–98.
- Surowiecki, James. 2005. *The Wisdom of Crowds*. Anchor.
- Tidwell, Lisa Collins, and Joseph B. Walther. 2006. Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time. *Human Communication Research* 28 (3):317–48.
- Traven, David. 2021. *Law and Sentiment in International Politics: Ethics, Emotions, and the Evolution of the Laws of War*. Cambridge University Press.
- Vertzberger, Yaacov Y.I. 1998. *Risk Taking and Decisionmaking: Foreign Military Interventions*. Stanford University Press.
- Voss, James F., and Timothy A. Post. 1988. On the Solving of Ill-structured Problems. In *The Nature of Expertise*, edited by Michelene H. Chi, Robert Glaser, and Marshall J. Farr, 261–85. Erlbaum.
- Walther, Joseph B. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research* 19 (1):52–90.
- Waltz, Kenneth N. 1979. *Theory of International Politics*. McGraw-Hill.
- Weeks, Jessica L.P. 2014. *Dictators at War and Peace*. Cornell University Press.

- Wendt, Alexander. 2004. The State as Person in International Theory. *Review of International Studies* 30 (2):289–316.
- Wheeler, Nicholas J., and Marcus Holmes. 2021. The Strength of Weak Bonds: Substituting Bodily Co-presence in Diplomatic Social Bonding. *European Journal of International Relations* 27 (3):730–52.
- Whyte, Glen. 1989. Groupthink Reconsidered. *Academy of Management Review* 14 (1):40–56.
- Yarhi-Milo, Keren. 2018. *Who Fights for Reputation in International Politics? Leaders, Resolve, and the Use of Force*. Princeton University Press.

Authors

Joshua D. Kertzer is Professor of Government at Harvard University. He can be reached at jkertzer@gov.harvard.edu.

Marcus Holmes is Associate Professor in the Department of Political Science at the College of William and Mary. He can be reached at mholmes@wm.edu.

Brad L. LeVeck is Associate Professor in the Department of Political Science at the University of California, Merced. He can be reached at bleveck@ucmerced.edu.

Carly Wayne is Assistant Professor in the Department of Political Science at Washington University in St. Louis. She can be reached at carlywayne@wustl.edu.

Acknowledgments

Thanks to Henry Atkins, Riley Carney, Brad DeWees, Emily Jackson, Austin Jordan, Max Kuhelj Bugaric, Daiana Lilo, Ethan Mallove, MD Mangini, Andras Molnar, Clay Oxford, Eric Parajon, Yon Soo Park, Heather Rodenberg, Andi Zhou, and the students in GOVT 204 and the Political Psychology and International Relations research lab at William and Mary for research and programming assistance. We're also grateful to Jack Levy, John Paschkewitz, Katy Powers, Ryan Powers, Ken Schultz, Lisa Troyer, Alan van Beek, Jessica Weeks, and audiences at ISPP, APSA, Washington University in St. Louis, the University of Pennsylvania, and the University of Georgia for feedback, as well as to Rose McDermott for enormously helpful conversations at the WCFIA security cluster conference in 2018, and to Keri Lemasters, Brooke Moore, Sarah Pollack, and Amy Stockton for making the study possible.

Funding

This research was funded by the Defense Advanced Research Projects Agency (award no. W911NF1920162). We acknowledge the support of the Weatherhead Center for International Affairs and the Institute for Quantitative Social Science at Harvard University.

Key Words

Political psychology; foreign policy decision making; aggregation problem; group processes; cognitive biases

Date received: April 26, 2021; Date accepted: November 30, 2021