The Democratic Peace and the Wisdom of Crowds

Brad L. LeVeck, University of California, Merced

Neil Narang, University of California, Santa Barbara

Forthcoming in International Studies Quarterly

Abstract:

The finding that there are few, if any, cases of war between democratic states has generated a great deal of interest. This paper proposes a new theory for the democratic peace that highlights a previously unexplored advantage that democracies have in crisis bargaining. We argue that because democracies typically include a larger number of decision-makers in the foreign policy process, they will produce fewer decision-making errors in situations of ultimatum bargaining. As a result, we expect that bargaining among larger groups of diverse decision-makers will fail less often. In order to test our hypothesis, we use experimental data where subjects engage in ultimatum bargaining games. We compare the performance of both individuals and small groups to larger groups of decision-makers. We find strong support for the idea that collective decision-making among larger groups of decision-makers decreases the likelihood of bargaining failure.

1. Introduction

There is hardly an empirical phenomenon in the field of international relations that has received the same level of academic attention as the democratic peace. The now well-established finding has two parts. First, and most famously, is that there are few, if any, clear cases of war between established democracies (Chan 1984; Kant [1795] 1969; Maoz and Abdolali 1989; Weede1984, 1992). Second, and somewhat more controversially, is that democracies are no less war-prone overall than other kinds of states (Bremer1992, 1993; Dixon 1993, 1994; Lake 1992; Small and Singer 1976). In other words, while democracies rarely – if ever – fight each other, their roughly equal rate of war participation on average means they are more frequently at war with nondemocratic states. This finding is of such potential importance to policymakers that has it has been subjected to numerous empirical checks in the last several decades and found to be surprisingly robust (Gartzke 1998, 2000; Kacowicz 1995; Lemke and Reed 1996; Maoz and Abdolali 1989; Maoz and Russett 1993; Oneal and Russett 1999b; Ray 1995; Rousseau et al. 1996; Russett 1993; Russett and Oneal 2001; Russett, Oneal, and Davis 1998; Signorino and Riter 1999; Small and Singer 1976; Thompson and Tucker 1997l; Dafoe 2011).[1] As Jack Levy notes "the absence of war between democratic states comes as close as anything we have to an empirical law in international relations"(1989, 270).

Perhaps not surprisingly, theories of the democratic peace have proliferated alongside empirical tests, in part because of the difficulty in accounting for the apparent dyadic nature of

[1] There may be thousands of books and articles on the democratic peace – too many to review here. See Rosato (2003) and Dafoe (2011) for more thorough reviews of the theoretical and empirical challenges to the democratic peace finding.

the observation. What is it about democratic institutions that facilitate peaceful relations among states that have them? For the most part, existing institutional theories examine two ways in which democratic institutions may affect the likelihood of war between states: first, democratic institutions can help align the interests of leaders with their citizens, and second, democratic institutions may change the quality of information conveyed by states during crisis bargaining.[2]

The first of these explanations is premised on the idea that democratic institutions are more likely to hold leaders accountable for the costs of war.[3] Although war can be an extremely costly and risky process for the citizens of a state, political leaders, who ultimately make the decision to wage war, rarely suffer these costs themselves. Instead, it is the individual citizens who pay the psychological and material costs of fighting in the form of lives lost and higher taxes. If leaders can expect to enjoy the benefits of winning a war with little to no exposure to the costs of war, they will be more inclined to fight a risky war rather than negotiate a compromise.

According to this view, representative forms of governments may better align the interests of the ruler with the ruled, by periodically holding leaders accountable to their citizenry (Doyle 1997, 24–25; Russett 1993, 38–39). Democracy thus makes leaders more sensitive to the costs of war and thereby decreases the probability that they would fight for personal gain (Maoz and Russett 1993; Russett 1996). If war is costlier for democratic leaders, they should be less willing to risk war on average compared to leaders of non-democratic states who can afford to

[2] For a survey of behavioral and normative theories of the democratic peace dating to Kant's Liberal Peace, see Rosato (2003) and Dafoe (2011). See Stevenson (2016) for a normative theory.

[3] See Rosato 2003 for a general review of the literature in support of this mechanism.

gamble with others' lives and resources. This heightened sensitivity to the costs of war may also explain why democracies fight with non-democracies more often. If democratic leaders are less willing to pay the cost of war, autocratic states should challenge democracies more frequently or make greater demands in the course of war bargaining, thereby increasing the risk of war.

A second popular explanation for the democratic peace focuses on how democratic institutions may influence crisis bargaining between states. Building off the bargaining model of war (Fearon 1995), this argument rests on the idea that if war is the result of bargaining failure due to private information or credible commitment problems, then something about democratic institutions must solve these problems such that they are more likely to find mutually beneficial bargains that avoid the costs of war. In particular, proponents of this argument suggest that democracies may be better able to resolve the informational problem that arises when sides have private information about their costs of war relative to the issues at stake (and the incentive to misrepresent these values to their opponent). For example, democratic decision-making processes are often more open and transparent, especially in cases where different representatives argue or negotiate over foreign policy in public forums (Schultz 1998; 2001). This greater transparency of democratic decision-making allows opposing states to better assess the true capabilities and resolve of democratic states (Schultz 1998).[4]

While both of these arguments suggest plausible mechanisms that might account for the democratic peace finding, neither one addresses the possibility that democracy may be associated with a superior foreign policy decision-making process. The first argument simply suggests that

[4] A related informational mechanism, domestic audience costs, has also received significant attention in the crisis bargaining literature. See Fearon (1994); Tomz (2007); Weeks (2008).

democratic institutions cause leaders to be more pacific by exposing them more directly to the costs of war. This should bias democracies towards peace in general, but does little to explain why – if democratic institutions heighten leaders' sensitivity to the costs of war which, in turn, causes non-democracies to exploit their pacific tendency to make greater demands – democracies do not perform worst on average in crisis bargaining situations (Bueno de Mesquita et al 1999). That is, there is no evidence that non-democratic states generally extract greater concessions from democratic states over time because they are more inclined to back down. The second argument incorporates our understanding of crisis bargaining by acknowledging that all parties – regardless of regime type – have an incentive to avoid war, but that democracies are better able to convey their own capabilities and resolve to opponents, and are thus less likely to be challenged in the first place when they are observed to be highly resolved. Yet, compelling as this argument is, it may be incomplete since the role the democratic decision-making process is strictly passive, allowing an opponent to better assess a democratic state's reservation price, but ascribing no distinct advantages to democratic foreign policy decision-making itself.

Drawing on a now well-established literature on the advantages of group decision-making, we propose a new theory for the democratic peace that highlights a previously underexplored advantage that democracies may have in crisis bargaining. Specifically, we argue that the diverse collection of independently-deciding individuals characteristic of democratic states is likely to produce fewer decision-making errors than individual leaders or even policy experts in situations of ultimatum bargaining. As a result of this collective intelligence, we expect that bargaining with a group of decision-makers will fail less often.

In order to test these expectations, we develop a simple experimental design that isolates one key difference between democratic and autocratic decision-making: democracies typically

have a larger group of decision-makers involved in the foreign policy process. Closely matching our experimental conditions with both the assumptions of the bargaining model of war and the "wisdom of the crowds" literature, we find strong support for the idea that collective decision-making decreases the likelihood of bargaining failure. Across experimental conditions, larger groups of decision-makers consistently outperform individuals in situations of ultimatum bargaining, whether they are matched against a smaller group of individuals (i.e. in a mixed-dyad) or other, similarly large groups. The findings suggest that existing theories of the democratic peace that appeal to shared normative values, accountability, or transparency may be correct, but also incomplete, as simply aggregating decision-makers' bargaining choices through a voting institution replicates two key features of the democratic peace finding in a controlled experimental set-up: democratic dyads avoid costly bargaining failure more than autocratic or mixed dyads, and democracies do no worse than other regime types in terms of bargaining outcomes.

2. Theory

2.1 The Wisdom of Crowds

In the opening anecdote of his popular book "Wisdom of the Crowds", James Surowiecki illustrates a classic example of how crowds may be wise. At a 1906 county fair in Plymouth England, British Scientist Francis Galton came across a weight judging competition in which members of a gathering crowd were lined up to place wagers on the weight of a fat ox. The best guess would win the prize. Seven hundred and eighty-seven diverse individuals (including expert butchers and farmers and non-expert clerks) tried their luck at guessing the ox's weight in an attempt to win prizes. When the contest was over, Galton borrowed the tickets from the

organization and analyzed the guesses, hoping to show that average voter was capable of very little. Adding the contestants' estimates together and calculating the mean, Galton used this number to represent the collective wisdom of the Plymouth crowd, acting as if the crowd were voting as a single person. Given the mixture of the crowd, which included relatively "smart" guesses from experts with relatively "dumb" guesses from non-experts, Galton undoubtedly expected the guesses would be way off. The crowd guessed the ox would weigh 1,197 pounds. The actual weight of the ox was 1,198 pounds. In Surowiecki's words, " the crowd's judgment was essentially perfect".

What Galton discovered in averaging the guesses of the Plymouth crowd was a phenomenon that has been reproduced in multiple real world and experimental settings: that under certain conditions groups of independent decision-makers can be remarkably smart, even smarter than the smartest members within that group. While it was certainly true that the "dumbest" members of the Plymouth crowd were considerably worst than the so-called "experts" as Galton predicted (each individual in the group was off by an average of nearly 55 pounds, with a standard deviation of roughly 62 pounds), their guesses were wrong in very different ways, with some individuals dramatically overestimating the weight of the ox and others dramatically underestimating its weight. In averaging a diverse set of individual guesses, the errors canceled out and thus produced a collectively wise decision. In other words, even if most people within a group are not particularly well informed or rational (lacking the ability and desire to make sophisticated cost benefit calculations), when those imperfect judgments are aggregated together, our collective intelligence is oftentimes superior to the smartest of decision makers (see Tetlock 2005).

The importance of this finding for studying the behavior of political and social groups was not lost on Galton. In particular, the analogy to a democracy where people of radically different abilities and interests each get one vote suggested itself immediately. In Galton's words: "The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes." (p. xii). Despite his own belief that power in society should belong to a select few with the best qualities for breeding, Galton later conceded that, "the result seems more creditable to the trustworthiness of a democratic judgment than might be expected." (p. xiii).

Not all crowds are wise, however. And, over time – as the implications of Galton's findings have been studied in various social contexts – a theory of collective intelligence has been gradually refined to include certain key criteria. Contemporary theorists emphasize that collective accuracy depends on a combination of both individual accuracy and diversity, and can be characterized by the simple mathematical identity below (Page 2008, Hong and Page 2004, 2009, 2012).

Collective accuracy = average accuracy + diversity.

"Average accuracy" in this equation refers to the average magnitude of each individual's error. "Diversity" refers to how different individual guesses are on average. What the first term in this simple equation makes clear is that crowds must know something about the issue at hand. If, individuals know nothing about an issue and are wildly wrong, then the crowd will still tend to be very wrong as well. After all, rockets are designed by groups of engineers, not lay-people. On the other hand, if a number of individuals do know something about the problem at hand, but are prone to making different types of errors, then aggregating their views can help make an accurate

decision because different errors will cancel one another out. As we discuss below, it is plausible that democratic decision makers are both accurate and diverse enough to give democracies an advantage in foreign policy decision-making.

In addition to these general rules, scholars in the psychology literature have also identified a number of specific conditions under which groups are unlikely to perform better (Cason and Mui 1997; Bone et al 1999; Rockenbach et al 2001; Cox and Hayne 2002; Puncochar and Fox 2004; Koch and Sutter 2005, Kerr, MacCoun and Kramer 1996).[5] For example, worse decision-making may emerge when designated leaders promote conformity and self-censorship, which can lead to "group-think" (Sniesek 1992, Kleindorfer et al. 1993; Mullen et al. 1994), Similarly, problems can arise when groups polarize the attitudinal judgments of their members (Davis 1992; Kerr et al, 1996; Cason and Mui 1997). Importantly, however, many of these conditions do not apply in our experimental setup; and, there are also good reasons to believe that democratic decision-making is less vulnerable to many of these harmful conditions. We describe these reasons in detail below.

2.2 The Wisdom of Crowds in Democracies versus Autocracies

If a diverse group of independently deciding individuals can be collectively wise, and this may be behind some of democracies' ability to formulate superior policy decisions, it is surprising that more attention has not been paid to this particular democratic advantage in foreign policy decision-making.[6] Perhaps democracies, by aggregating predictions from a diverse

[5] In the interest of space, we review the results of these papers in the Appendix.

[6] One exception is an important study by Reiter and Stam (2002), who apply a similar logic to a different empirical puzzle: why democracies win the wars they initiate. Reiter and Stam argue

population of intelligent agents may outperform a team comprised of even the best-performing agents. That is, it might be the case that democracies have an advantage in foreign policy decision-making when compared against alternative institutional forms like autocracies that aggregate information from a smaller, less diverse set of "expert" individuals.

Even though foreign policy decision-making in democracies is oftentimes dominated by a relatively small group of educated elites (Saunders 2011, Hafner-Burton, Hughes, and Victor 2013), there are still compelling reasons to believe that democracies draw on a larger, more diverse set of views *on average* when making decisions about war-bargaining. First, by holding periodic elections, citizens can express their views on which leader or mix of representatives is best suited to conduct international affairs. Indeed, existing evidence suggests that citizens, while hardly being experts in foreign policy, do hold broadly informed opinions on such matters, see clear differences between the candidates on issues of foreign policy, and vote partially on the basis of these factors (Aldrich 1999). Citizens may therefore elect representatives who take a particular approach to foreign policy, such as whether a state should take a more hawkish or dovish approach to matters of interstate conflict (DeNardo 1995), but leave the details of how to

---

that democracies "are better at forecasting war outcomes and associated costs" because they "benefit from more and higher quality information" (p.23), and thus only initiate winnable wars. They argue, "the unitary nature of dictatorships…forgoes democratic advantages from the market-place of ideas that provide broad checks on a single leader" (p. 25). Reiter and Stam build from Schultz (1999), who also raises the prospect that democracies are more strategic about what conflicts they enter. Here, we explore whether this advantage helps democracies forecast the reservation price of opponents in crisis bargaining, and whether it offers a partial explanation for the democratic peace (i.e bargaining success, rather than war outcomes).

best implement a given approach to elected representatives and the bureaucrats they oversee (Lupia and McCubbins 1994, 2003). The diverse approaches of different elected officials (many of whom have some input into the foreign policy decision-making process) may act like the diverse heuristics and interpretations found in recent models of collective wisdom (Hong and Page 2004, 2009, 2011). Second, citizens in democracies can more efficiently express approval or disapproval for their leader's policies through public polls. Again, these polls may aggregate citizens' diverse views on the wisdom of a particular approach to foreign policy. Third, democracies tend to have freer markets with exchanges that can react almost instantly to inform leaders about the expected outcome of a particular policy choice (Gartzke 2007, Wolfers and Zitzewitz 2009). These market signals can act like weighted votes from market investors. Finally, democracies tend to establish different domestic institutions with diverse approaches or perspectives on foreign policy. For instance, in the US, the Departments of State and Defense have different intelligence sources, decision-making structures, and personnel.[7] Yet, both institutions may have input on how to deal with a particular adversary.

Together, these information aggregation mechanisms allow for more diverse groups of independently deciding individuals to process information separately and express their own independent assessment on foreign policy matters. Thus, even though the decision to go to war in a democracy like United States may ultimately rest with only a small group of leaders gathered in a "situation room", existing studies support the comparative-static claim that democratic decision making is – on average – relatively more pluralistic than autocratic decision-making due

[7] In other words, even though cabinet members' views may be correlated by a shared ideology or by a desire to gain favor with an ideological leader (Saunders 2011), in many contexts, ideology will not induce perfect correlation.

to these mechanisms of accountability. Furthermore – even when aggregating similar beliefs across similar numbers of individuals – participants in autocracies often lack the incentive to tell leaders the truth (Reiter and Stam 2002). And although elites may often influence or manipulate the preferences of citizens in democracies (challenging the assumption of independence) (Zaller 1992; Lenz 2012), existing studies suggest that democratic decision-making is influenced by a more diverse set of opinions on average relative to autocratic states.[8]

Even at the level of elite decision-making—outside the direct influence of everyday citizens—there is little controversy in the academic literature that democracies tend to have a larger group of decision-makers involved in the foreign policy process. At the broadest level, the Polity IV index measure – on which the democratic peace phenomenon is based – is primarily driven by the variable *XCONST* (Gleditsch and Ward 1997), which, in large part, codes the number of actors across institutions that constrain policymaking by the executive. The variable therefore reflects the fact that democratic policy making is typically influenced by a larger number of independent actors. Similarly, the POLCONIII (Henisz 2000) index used in some robustness checks of the democratic peace (Tsebelis and Choi 2009) measures the raw number of institutional veto players and their relative independence in terms of preferences and ideological

[8] For example, consider that even when partisan media, like Fox News or MSNBC, heavily influences citizens' views, (1) even these opposing views are likely to create diversity in opinion with errors that cancel out, and (2) some component of citizens' opinions still remains statistically independent (i.e. unexplained) by these "elite" opinions (Levendusky 2009). The experiment below can be understood to capture this independent component.

view points.[9] As we review further in the Appendix, there is also evidence that these veto players have some influence over foreign policy, not just domestic policy.

There is also plenty of qualitative evidence to support the assumption that democracies contain a larger, more diverse group of individual decision makers on average. For example, in categorizing foreign policy decision making across states over time, Hermann and Hermann (1989) show that autocratic regimes are almost perfectly correlated with "Predominant Leader" or "Single Group" decision units that "will be relatively insensitive to discrepant advice and data" (365), while foreign policy-making in democratic regimes is correlated with "Multiple Autonomous Actors".[10]

Even in the United States, where the executive branch is thought to enjoy a great deal of autonomy – particularly over decisions to go to war – there nevertheless exists a robust and well-documented interagency process as a mechanism for collective decision-making. At multiple levels, the U.S. interagency process draws together a diverse collection of independently-deciding actors from across multiple agencies with distinct – sometimes

[9] In new Appendix Table A6, we compare democracies and autocracies along both variables quantitatively and show that democracies are systematically characterized by a larger, more diverse group of independently deciding individuals on average.

[10] Geddes (2009; 2013) and Weeks (2012; 2014) have also detailed intricate decision-making processes across different types of autocratic regimes.

parochial, often times conflicting – interests and beliefs based on independent characterizations of the international system (Raach et al. 1995; Marcella 2004; Gorman et al. 2005).[11]

Detailed historical accounts illustrate how this interagency process can aggregate a large and diverse number of views. In his seminal article "Conceptual Models and the Cuban Missile Crisis", Allison (1969) provides what is perhaps the most well known example of how US foreign policy outputs are "the consequences of innumerable and oftentimes conflicting smaller actions by individuals at various levels of bureaucratic organizations in service of a variety of only partially compatible conceptions of national goals, organizational goals and political objectives" (63). Specifically, Allison shows that Kennedy struggled to weigh different, and sometimes conflicting, recommendations from his closest advisors drawn from different agencies with different perspectives. The moves were "resultant of collegial bargaining" (691) from a "conglomerate of semi-feudal, loosely allied organizations, each with a substantial life of its own"(698). As Allison notes, "the nature of problems of foreign policy permits fundamental disagreement among reasonable men concerning what ought to be done. Analyses yield conflicting recommendations. Separate responsibilities laid on the shoulder of individual personalities encourage differences in perceptions and priorities…More often, however, different groups pulling in different directions yield a resultant distinct from what anyone intended" (707).

[11] Indeed, despite the presence of a dedicated Intelligence Community, organizations in the US Federal Government maintain their own intelligence agencies precisely to arrive at independent assessments and avoid "group-think": DoD operates the Defense Intelligence Agency, State operates the Bureau of Intelligence Research, Treasury operates the Office of Intelligence Analysis, etc.

In the US government, these actors include "Chiefs": the President, Secretaries of State, Defense, and Treasury, Director of the CIA, Joint Chiefs of Staff and, since 1991, the Special Assistant for National Security Affairs" (709).

Allison's account of the decision to implement a blockade of Cuba during the crisis provides an excellent illustration of how inputs from numerous, diverse viewpoints – even from within the executive branch, where members often have a shared ideology (Saunders 2011) – can have a significant impact on crisis bargaining. As described by Allison, Senators Keating, Goldwater, Capehart, Thurmon and others initially attacked Kennedy for his "do nothing approach", while McGeorge Bundy, the President's Assistant for National Security Affairs, asserted that there was no present evidence that the Cuban and Soviet Government would attempt to install a major offensive capability (712). Meanwhile, Colonel Wright and others at DIA believed that the Soviet Union was placing missiles in Cuba – information that fell on the diverse crowd of advisers differently (713). Kennedy's principal advisors, including Secretary of Defense McNamara, McGeroge Bundy, Theodore Sorenson, and the President's brother Robert Kennedy, considered two tracks: do nothing and taking diplomatic action (714). However, the Joint Chiefs of Staff believed that the time was clear for a military invasion of Cuba (Allison 1969; 714). According to Allison, "the process by which the blockade emerged is a story of the most subtle and intricate probing, pulling and hauling; leading guiding and spurring". Initially, Allison notes, "the President and most of his advisers wanted the clean, surgical air strike" (714). Remarkably, however, despite the presence of sizeable minority preferring an air strike, the President ultimately opted for a blockade after considering the advice of McNamara and Robert Kennedy (714). Reflecting on the influence of the diverse opinions of his advisors, the President's brother claimed that "the fourteen people involved were very significant" (714).

In stark contrast to the Kennedy administration's handling of the Cuban missile crisis, the overwhelming consensus among diplomatic historians on the Cuban Missile Crisis is that Kennedy's counterpart in the Cuban Missile Crisis, the Soviet leader Nikita Khrushchev, drew from a much smaller group of advisors than Kennedy, and that those advisors who Khrushchev did consult with were systematically ignored during the Crisis, if they even felt safe to express their true beliefs at all (Fursenko and Naftali 1998; Fursenko and Naftali 2007; Taubman 2003; Dobbs 2008). Beyond the case of the Cuban Missile Crisis, Hermann et al. (1989) use four case studies to demonstrate how autocratic regimes made the decision to initiate or escalate war after periods of failed negotiations due to their relative insensitivity to discrepant advice and data. In a more recent example, Saddam Hussein repeatedly ignore the advice of his military advisors and scientists (many of whom were afraid to express dissent in the first place), many of whom correctly estimated that the rate of Iraq's nuclear program ran a high risk of triggering war (Braut-Hegghammer 2016). This further illustrates how autocracies may be worse at incorporating knowledge dispersed among multiple actors, even when those actors hold key advisory roles in government.

2.3 The Wisdom of Crowds and the Democratic Peace

The possibility that a more diverse collection of independently-deciding individuals characteristic of democratic states might be superior to non-democracies in predictive tasks has important implications for the democratic peace finding. As noted above, the prevailing wisdom holds that crisis bargaining with democracies fails less often because they are better able to convey private information about their *own* capabilities and resolve due to their greater transparency. However, based on the logic outlined here, we propose an alternative mechanism through which democracies may be able to resolve the informational problems that lead to

bargaining failure. For the reasons outlined above, we posit that democracies are better able to aggregate and interpret noisy signals gathered during a crisis in way that cancels out decision-making errors.

Consider the simplest model of crisis bargaining as outlined by Fearon (1995). In this setup two states (S_1 and S_2) have divergent preferences over the division of some issue space represented by the interval $X = [0,1]$, where each state's utility normalized to a 0,1 utility space. S_1 prefers issue resolutions closer to 1, while S_2 prefers resolutions closer to 0. Supposing states fight a war, S_1 prevails with probability $p \in [0,1]$ and gets to choose its favorite outcome closer to 1. S_1's expected utility is $p\mathrm{u}_1(1) + (1 - p)\mathrm{u}_1(0) - c_1$, or $p - c_1$. S_2's expected utility for war is $1 + p - c_2$. The parameters, c_1 and c_2 represent the costs for fighting a war to each side along with the value of winning and losing on the issues at stake. Importantly, the costs of fighting open up a range of bargained solutions between each states' reservation price, $p - c_1$ and $p + c_2$, that both sides should strictly prefer to paying the costs of war. Structured this way, the puzzle becomes why sides ever fail to identify a negotiated settlement within this range ex ante, knowing that war is always inefficient ex post.

Fearon suggests that coherent rationalist explanations for war will fall into one of two categories: sides can fail to reach a bargain because they have private information with incentives to misrepresent or because sides are unable to credibly commit themselves to follow through on the terms of the agreement. According to the first explanation, sides have asymmetric information about their own capabilities, p, and resolve, c, and they have an incentive to over-represent (or under-represent) their ability on these dimensions to their opponent in order to secure a better settlement.[12] As a result, while the costs of fighting open up a range of negotiated

[12] While Fearon discusses private information by bluffing about greater strength, later research discusses private information by feigning weakness (Slantchev 2010).

settlements both sides prefer to war, the incentive to bluff may lead sides to delay settlement in favor of fighting in order to accrue enough information to formulate reliable beliefs about their opponent's strength (Slantchev 2004).

In situations of incomplete information, war (bargaining failure) can occur in Fearon's model if State 1 overestimates State 2's cost of going to war, and therefore makes an offer that is too small for State 2 to accept. On the other side of the decision, war can also occur if State 2 underestimates its own costs of war, and chooses to only accept offers that State 1 would not reasonably propose. In each of these cases, decision errors can happen because decision-makers have uncertainty about key parameters, and they can only estimate these parameters with some error. However, it is possible that the error made by one decision-maker within a state may be different from that of another. For example, while one decision-maker might over-estimate the other State's cost of going to war, another decision-maker could err in the opposite direction. If such views are aggregated, the errors could cancel out.

In the next section, we describe a version of the classic ultimatum game, and we use this model as the basis for an experimental research design in which we test the proposition that regimes with more decision-makers experience fewer instances of costly bargaining failure (analogous to war), and achieve outcomes that are at least as good as the outcomes achieved by regimes with fewer decision-makers.

3. Methodology and Results

Using observational data to identify the effect of information aggregation mechanisms on war bargaining outcomes is difficult for a number of reasons. First asymmetric information

presents the same problem for the analyst that it does for states in the international system: a state's reservation price for war is private information that is rarely revealed. This makes it difficult to know how close one state's offers are to another state's reservation price for costly conflict. This is especially true for the majority of crisis bargaining scenarios, because offers rarely trigger war. Even in the rare cases where crisis bargaining devolves into war, it is impossible to know with any certainty just how much one state's offer fell short of another state's threshold for avoiding conflict.

Second, in an uncontrolled environment, it is difficult to ascertain what information individual decision makers had access to, and exactly how that information was filtered through executive decision-making processes. While it is important for future work to trace the precise process by which signals about opponents are aggregated, and how these aggregated signals influence state decision-makers, this approach is not ideal for clearly answering the more primary question of whether aggregation *can* influence bargaining in the manner predicted by existing theories. Such questions are better answered in an environment where the researcher can carefully control what information actors have access to, and how that information is aggregated.

3.1 An Experiment

To examine the question of whether information aggregation can improve bargaining outcomes we look at data from laboratory bargaining games. Specifically, we look at a variant of the Ultimatum Game (Guth 1982), which (as we further explain below) mimics key features of war bargaining.[13] The game is played between two players, a *proposer* and a *responder*, who

[13] We use the Ultimatum Game instead of the games used by Tingley and Wang (2010) and Tingley and Walter (2011), which allow the experimenter to manipulate responders' cost of

bargain over a fixed pie of 100 monetary units (mu). The proposer makes an integer offer, $S_p \in [0,100]$, which is the portion of the pie she proposes keeping for herself. The responder simultaneously makes a demand, $S_r \in [0,100]$, which is the minimum portion of the pie they will accept without rejecting the proposer's offer. The monetary payoffs for the proposer and responder are:

$$\begin{array}{cc} (S_p, 100 - S_p) & \text{if } 100\text{-}S_p \geq S_r \\ (0,0) & \text{if } 100\text{-}S_p < S_r \end{array}$$

In other words, if the proposer's offer exceeds or equals the responder's demand, then the pie is split according to the proposer's offer. If the offer falls short of the demand then the offer is rejected, and both parties receive 0 mu.

If proposers' and responders' utility is strictly increasing in the amount of money they personally receive – and they both have mutual knowledge of this fact – then the unique subgame perfect Nash equilibrium for the Ultimatum Game is for Proposers to offer zero, and for Responders to accept zero because they are indifferent between accepting and rejecting. If this

---

bargaining failure. We did this for two practical reasons: First, compared to the laboratory, it is more difficult to ensure that subjects in online experiments fully understand complex instructions (Rand 2012, 176). We therefore chose the Ultimatum Game, in part, because it was the simplest game that met our requirements. Second, there have now been hundreds of experiments conducted using the Ultimatum Game, including international policy elites. We could therefore examine how well crowds performed relative to individual experts.

theoretical expectation were to hold, this might make the Ultimatum Game a poor analogy to the bargaining model of war because only the proposer is strictly worse off when an offer of 0 is made and rejected. However, the existence of this strategy profile does not present a major problem for testing our theory because, as a practical matter, individuals in the Ultimatum Game almost never propose 0 or set 0 as their minimum acceptable offer across real world settings (Camerer 2003). Thus, empirically, these potential offers – while theoretically possible – have no practical effect on our results below.[14]

The infrequency of proposals that offer zero in the Ultimatum Game is likely due to the fact that responders exhibit aspects of real world bargaining that are crucial for our particular question: they have positive but variable minimum acceptable offers (Camerer 2003; Henrich et al. 2001). This is because subjects derive utility from other things besides monetary payoffs – like satisfying norms of fairness or feelings of spite. So while the responder cannot possibly gain a higher payoff by demanding more, this is only true in terms of *monetary* payoffs. In terms of players' *utility* for monetary splits, things are often different. This means that responders can rationally demand more than zero, and proposers can anticipate this by offering some positive amount to avoid bargaining failure. Numerous experiments have shown that responders varied thresholds for rejecting an offer do not purely reflect a mistake, but rather some actual differences in players' utility for different monetary splits (Camerer 2003, Andreoni and Blanchard 2006).

[14] Indeed, individuals in our experiment vote to propose 0 just over 4% of the time, but, in most cases, these votes do not manifest in observing a proposal of 0 because the votes occurred as part of a group in which votes for larger proposals bring the actual observed frequency of proposals that offer zero to substantially less than 1%.

Crucially, heterogeneity in demands creates uncertainty for proposers regarding what offers will and will not trigger costly bargaining failure. In this regard, the experiment is analogous to many models of war bargaining under asymmetric information, such as Fearon (1995) or Powell (1999), where the proposer makes a single take it or leave it offer under uncertainty about an opponent's costs of war (i.e. opponent-type). Such decision-making errors are analogous to a leader underestimating its opponent's willingness to fight. Rejection in our game is analogous to a costly outside option, such as war, which both players wish to avoid in favor of some mutually acceptable bargain.

While the ultimatum game is a workhorse of laboratory studies on bargaining, our innovation is to systematically manipulate the number of decision-makers on each side, and see how this affects the rate of costly bargaining failure. Other articles have looked at what happens when subjects' views on how to play the Ultimatum Game are aggregated by deliberation (Bornstein and Yaniv 1998) and voting (Elbittar, Gomberg, and Sour 2011). However, no study to date has examined what happens to the rate of bargaining success when the number of decision-makers on each side is systematically varied. Our experiment does this with respect to voting, which is a common way for aggregating decisions.

Even though previous studies of individual bargaining in the Ultimatum Game suggest that decision-makers avoid bargaining failure a large fraction of the time (Camerer 2003), it is far from guaranteed that aggregating subjects' views will further increase the proportion of successful bargains in a population. For one, subjects may have informed views about how to bargain with other individuals, but may be relatively uninformed when it comes to bargaining with groups of different sizes. Second, the size of a group itself may diminish individual decision-makers' incentive to make wise decisions (Downs 1957). Making a wise vote takes

mental-effort, but that effort can be potentially rendered moot by other voters' decisions (Downs 1957; Popkin 1991). Furthermore, simply knowing that you are part of a group may make one more aggressive towards other "out groups", such as the group you are bargaining with (Tajfel and Turner 1979); and this aggression might plausibly lead to increased bargaining failure. Whether these potential pitfalls of collective decision-making can be overcome by its advantages is an empirical question, which we test.

> *Our hypothesis is that decisions aggregated from larger groups of proposers and responders will lead to fewer instances of bargaining failure and higher earnings compared to smaller groups and individuals.*

To test this, we modified an experiment by Rand et al. (2013), where proposers and responders were asked to play a single round of the Ultimatum Game described above.[15] In the original experiment, each proposer submitted a single offer while each responder submitted a single demand simultaneously. Demands and offers were then paired at random, and subjects were paid accordingly. Thus, each proposer had an incentive to make a proposal that would yield the highest expected earnings when played against a random (anonymous) responder. The expected success of each proposer's offer in the experiment can be calculated based on how often it would have been rejected by the population of proposers, and how many monetary units each proposal would have earned on average.

In our modification to this experiment, we compare the success of offers and demands made by small groups of 3 individuals to the success of offers and demands made by much larger groups of 9 individuals. These smaller groups of size 3 in the experiment are analogous to

[15] It is possible that crowds might have additional advantages that would emerge in a more dynamic setting. Future experiments might explore group advantages in learning.

autocracies, which tend to have a smaller number of decision-makers included in the policy-making process. Larger groups of size 9 are taken as analogous to more democratic polities, where more individuals are typically involved in the policy-making process. We use a group size of 3 for autocracies because it is the smallest size that has a well-defined majority. Henceforth, we refer to small groups as "Autocracy" and large groups as "Democracy". Of course, all the caveats with this stylized operationalization still apply (see section 2.2 and section 4). We use a group size of 9 because it was one of the largest treatment "dosages" we could implement while still having enough observations to test our directional hypothesis (that larger groups of decision-makers decrease the rate of bargaining failure). However, in new Appendix Figure 1, we test whether our results are particularly sensitive to using 9 players (as opposed to smaller groups 5 or 7). We find evidence that our results are robust to these differences.

A group's proposal to the other side was determined in the following manner: each individual in a group simultaneously and anonymously submitted a vote for what their group should offer to the other side. The group's actual proposal was then the median offer submitted in the group. For example, say that in a group of 3, individuals voted to offer 17, 18 and 24. The group's actual offer would be 18. While this procedure certainly does not capture the intricacies of foreign policy decision-making in a democracy or any other state, it is akin to a decision rule where the median voter's preference is decisive, and thus it approximates a number of real-world collective decision-making bodies, such as voting in elections (Downes 1951) or Congress (Krehbiel 1998). Specifically, aggregation processes like this one can be understood as similar to citizens voting for politicians with a particular level of hawkishness or dovishness, representation across bureaucracies in interagency meetings (Allison 1969, Janis 1972), or congressional votes over war authorization/war funding during crisis bargaining. While there are many significant

differences across each of these aggregation mechanisms, they all collect a large number of diverse viewpoints and aggregate them into a single number or outcome that can influence or determine foreign policy.

Of course, the downside of our stylized procedure is that it abstracts away from the intricacies of any one of these mechanisms. However, the upside is that it captures our key independent variable in a way that is tractable and relatively easy to interpret. We further discuss concerns over the external validity of this mechanism in a subsequent section below.

It is also worth noting that, in the absence of deliberation, "groupness" in our experiment emerges from individuals being informed about whether or not they were playing in a group before making their votes. Thus, individuals cast their vote in expectation of it being aggregated. Therefore, our treatment induced any behavioral changes that would arise from subjects *knowingly* voting as part of a group to influence the final proposal. And despite the presence of deliberation in the real world (and the attendant risk of attenuating the wisdom of the crowds), our discussion above illustrates that the risk of "group think" from deliberation is much more severe in autocracies, where "Predominant Leader" or "Single Group" decision units are "relatively insensitive to discrepant advice and data" Hermann and Hermann (1989). Therefore, while our voting mechanism does not fully capture some of the dynamics that might emerge from deliberation, it does preserve the fact that democratic deliberation typically involves a larger number of more independent inputs.

We posted this experiment online and recruited 1409 subjects through the internet labor market Amazon Mechanical Turk.[16] Subjects were paid $0.50 as a show-up fee simply for

[16] See the supplementary appendix for further details on our recruitment procedure.

participating in the experiment. Subjects were randomly assigned to be on Side A or Side B. Players were told that Side A's task was to propose to Side B how much of $0.40 cents should go to each member of Side B and how much should go to each member of Side A. For example, each member of Side B might get $0.10, implying that each member of Side A would get $0.30[17]. Side B would decide what minimum amount was an acceptable offer. If Side A's offer to Side B was at or above Side B's minimum acceptable offer, then both players were paid the bonuses according to the proposed division. Otherwise, no member of either side earned a bonus.

We defined the total size of the pie in terms of what each member received, so that the individual stakes of the decision were constant across conditions. In other words, changing the group size across conditions did not change the absolute amount of a fixed prize that each individual in a group could receive. While this decision was made primarily to improve the experiment's internal validity (by isolating the effect of aggregation rather than an individual's stake in the decision), it does have a real world analogue. Whereas the benefits of any bargain are typically more diffuse in large populations when the stakes are strictly material, there are many conflicts where one polity might impose a different "way of life" on citizens in another country (Lake 1992). In these situations, citizens and other decision-makers might place the same value on their own "way of life" regardless of how many other citizens exist in the country.

To ensure comparability of our study to existing studies, we began by first randomly assigning 232 of the subjects (out of 1409) to a baseline condition of a single proposer making a

[17] The size of the pie is always shown as $0.40. We used numerical examples in the instructions to illustrate how the $0.40 would be divided as a result of the proposal, but the hypothetical payoffs used were drawn randomly, so as not to systematically bias players' strategies.

take-it-or-leave-it offer to a single responder (the canonical ultimatum game). We then randomly assigned each of the remaining 1,177 subjects to one of our four experimental conditions:

1. A small group of 3 proposers making a take-it-or-leave-it offer to a small group of 3 responders (Autocracy/Autocracy);
2. A small group of 3 proposers making a take-it-or-leave-it offer to a large group of 9 responders (Autocracy/Democracy);
3. A large group of 9 proposers making a take-it-or-leave-it offer to a small group of 3 responders (Democracy/Autocracy);
4. A large group of 9 proposers making a take-it-or-leave-it offer to another large group of 9 responders (Democracy/Democracy)

Subjects were informed that the voting mechanism for group decision-making would simply be the highest offer for which there was a majority support, as described above. A summary of the conditions is shown below in Table 1.

Table 1: Four Ultimatum Bargaining Experimental Conditions

Side B

	Autocracy (3 Responders)	**Democracy (9 Responders)**
Autocracy (3 Proposers)	***Condition 1 (N= 124, 110)***	***Condition 2 (N=85, 280)***

	Democracy (9 Proposers)	***Condition 3 (N= 286, 98)***	***Condition 4 (N= 92, 102)***
Side A			

For each of our experimental conditions, we estimated how well each side would do on average, both in terms of avoiding bargaining failure and in terms of how much individuals earned, by randomly drawing 1,000 samples (with replacement) of *k* group members from the *N* subjects who participated in that experimental condition. For instance, in the Democracy/Democracy condition, we would randomly draw a set of 9 proposers out of all the subjects in the pool who were assigned to this condition, and another set of 9 responders who were assigned to this condition. We would then measure whether bargaining succeeded or failed by whether proposers collectively made an offer that was greater to or equal to what the responders collectively demanded. To obtain standard errors for this estimator, we used the non-parametric bootstrap, running our procedure over 3000 samples of the data.

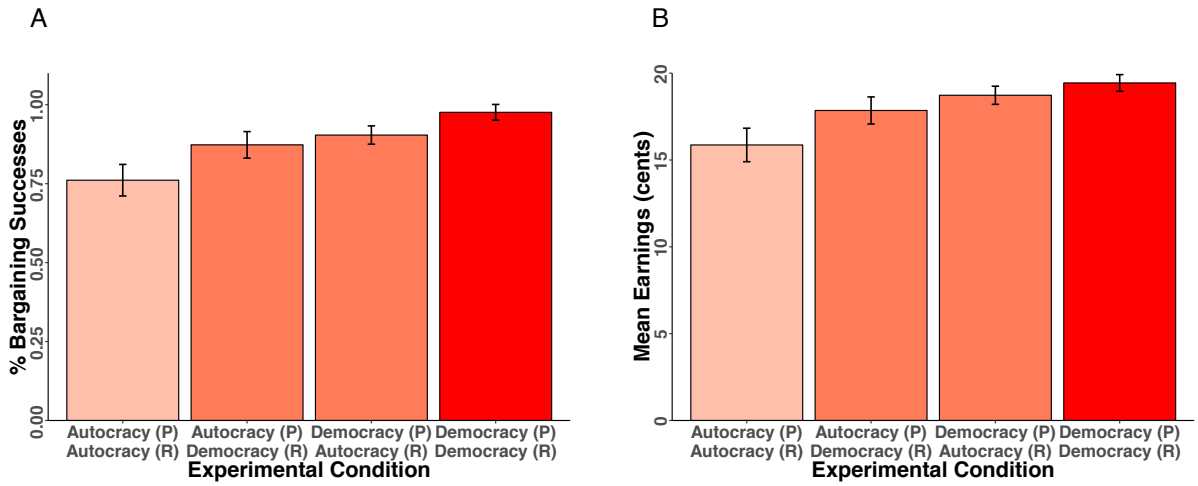
3.2 Results

We began by confirming that we could replicate past studies of one-on-one bargaining between individuals in the Ultimatum Game using the 232 subjects in our Baseline Condition. Similar to past studies, our results show that individuals avoid bargaining failure approximately 75-percent of the time (Camerer 2003). Specifically, individuals in this Baseline Condition of our experiment avoided bargaining failure 76.5 percent of the time (95% CI [0.70 to 0.83]).

Next we examined each of our main experimental conditions. **Figure 1** shows the estimated mean outcome in each condition, with bootstrapped standard errors from 3000 sub-

samples of the data. Moving from left to right along the X-axis are the four experimental conditions. Condition 1 is labeled "Autocracy/Autocracy", Condition 2 is labeled "Autocracy/Democracy", Condition 3 is labeled "Democracy/Autocracy", and Condition 4 is labeled "Democracy/Democracy".

Figure 1: Bargaining Failure and Earnings Across Treatments



In **Panel A** of **Figure 1**, the Y-axis represents the percentage of times bargaining succeeded, or – in our analogy – the costly reversion outcome of war was avoided. In **Panel B**, the Y-axis represents the average earnings of proposers in each condition. We investigated players' earnings to distinguish our hypothesis that groups in situations of ultimatum bargaining are collectively wise (by making more efficient proposals that more closely predict the reservation price of their opponent) from the alternative possibility that groups exhibit a lower rejection rate simply because they bargain in a way that is more risk-averse and inefficient (with groups consistently offering more generous proposals in order to secure a peaceful settlement at any cost).

Beginning with the Autocracy/Autocracy condition at the far left of Panel A, our results show that small groups of 3 do no better with respect to the percentage of times bargaining succeeds compared to the Baseline Condition described above, in which individuals faced individuals and bargaining succeeded roughly 75-percent of the time, (76.1, 95% CI [0.70 to 0.83]). Consistent with the Wisdom of the Crowds hypothesis, however, we find that mixed dyads, in which even one side is a large group of 9, performs significantly better in situations of ultimatum bargaining compared to dyads with two small groups. Autocracy/Democracy dyads avoid conflict 87.3-percent of time (95% CI [0.79 to 0.96]), and Democracy/Autocracy dyads avoid conflict 90.4-percent of time (95% CI [0.85 to 0.96]). Also consistent with our theory, democratic dyads perform the best, avoiding bargaining failure 96.7-percent of time (95% CI [0.93 to 1.00]). In other words, ultimatum bargaining between Democracies rarely if ever fails.

In **Panel B**, we investigate earnings across the four conditions for the reasons outlined above. These findings mirror the result in Panel A, with mixed dyads earning significantly more than autocratic dyads, and democratic dyads earning more than even mixed dyads on average. Democratic dyads earned on average 19.4 cents compared to Autocratic dyads in which individuals earn 15.9 cents on average. This suggests that proposals of large groups are better calibrated to the demands of responders, which is consistent with the hypothesis that democracies are "wiser", and is also consistent with the finding in observational studies that democracies do not perform worst on average in crisis bargaining situations (Bueno de Mesquita et al 1999). These higher earnings do not emerge because larger groups, on average, make substantially more generous offers. Instead, higher earnings emerge because aggregation averages out overly aggressive offers from individuals that would normally trigger bargaining

failure, and also offers that would be far too generous.[18]

3.3 Why is the Result Not Strictly Dyadic?

The results above clearly replicate the important dyadic aspect of the democratic peace finding: democracies almost never fight each other. However, it is not obvious from Figure whether our results replicate the more controversial finding that democracies are no less war prone overall, which implies that mixed dyads should be more war prone than even autocratic dyads (Gleditsch and Hegre 1997).[19] In the online appendix, we discuss two potential reasons why decision-aggregation may appear to have a monotonic effect in our experiment, but a dyadic effect in the real world. First, mixed dyads may have an overall higher rate of dispute *initiation* that fully offsets the benefits of aggregation within a crisis. Second, factors not present in our experiment could lead the different types in mixed dyads to have systematically biased views about how to bargain with another type, and this could cause aggregation to actually produce worse bargaining outcomes in mixed dyads.

3.4 Additional Tests

[18] The median offer from autocracies and democracies was both 20 and the mean was both 17. If we condition on bargaining success, democracies and autocracies earn roughly the same amount in our experiment. This replicates other findings in the literature, which suggest that democracies do not do appreciably worse in the bargains they successfully conclude short of war (Bueno de Mesquita et al 1999).

[19] See Gleditsch and Hegre (1997) for a summary of the controversy over, and mixed results for, a monadic democratic peace.

A second aspect of the Wisdom of the Crowds hypothesis is that crowds of individuals can even outperform expert individuals in predictive tasks (Tetlock 2005). Above, we discussed the possibility that democracies, by aggregating predictions from a larger number of decision-makers, may outperform even relatively skilled experts in bargaining scenarios that mimic key aspect of war-bargaining. To investigate this, we compared the performance of democratic dyads in our experiment to three types of individuals. The first type are 'Inexperienced Individuals". These are individuals from our Baseline Condition who, in a post-experiment survey, reported that they had never played a game similar to our ultimatum game scenario.[20] The second type of individuals that we compared to democratic dyads is "Experienced Individuals", who reported that they had played a similar game in the past (50-percent of the subjects in our Baseline Condition). The third type of individuals is "International Policy Elites". This is a sample of 102 international foreign policy elites who were recruited to play an Ultimatum game in a previous study by LeVeck et al. (2014). These elites had significant real world experience in actual international bargaining.

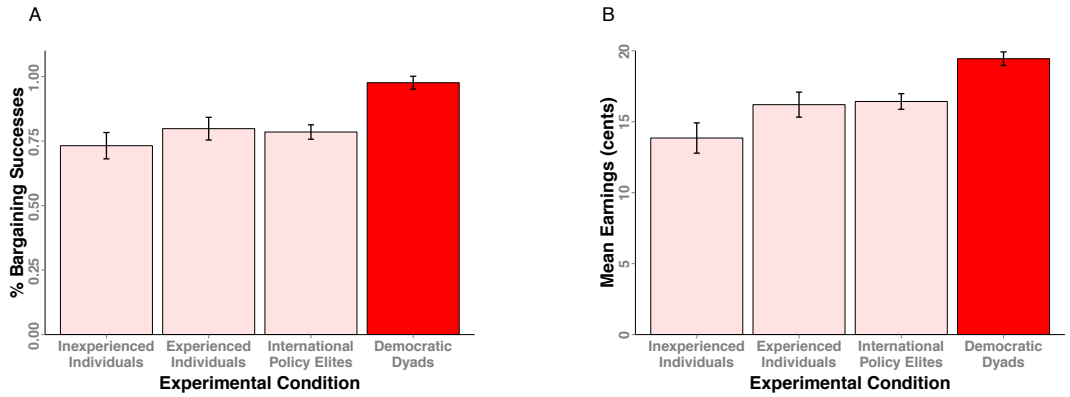
Figure 2 compares the results of each type of individual against the performance of democratic dyads along the same two dimensions. Panel A shows the percentage of time bargaining succeeded, and **Panel B** shows the average earnings of proposers in each condition.[21]

[20] Specifically, Inexperienced Individuals were subjects who did not answer "yes" to the following post-experiment question: "Have you ever played a similar game, where one player proposes how to split a monetary prize and another player decides whether to accept or reject the offer?"

[21] The elite sample from LeVeck et al. (2014) played for a larger monetary prize. We have therefore rescaled earnings to match the prize used in our study.

Beginning in Panel A, the results show that experienced individuals and international policy elites avoid bargaining failure more than inexperienced individuals, though this difference is not statistically significant at conventional levels. At the same time, the results in Panel B show that both groups of expert individuals earn significantly more than inexperienced individuals. Meanwhile, the results in both panels strongly confirm the Wisdom of the Crowds hypothesis. Democratic dyads comprised of both experienced and inexperienced individuals, dramatically outperforms even experts on both measures. These results are consistent with the findings of Tetlock (2005).

Figure 2: Group vs. Expert Performance



Finally, a third aspect we investigated is what factors are actually driving the observed behavior in our experiment. We do this because our aggregation mechanism may actually aggregate two distinct factors: behavioral norms and knowledge about what the other side's minimum acceptable offer will be (Camerer 2003). Because our theory focuses on the second element, beliefs, we isolated that component to see if our main hypothesis holds. In the Appendix Figures A2 and A3, we show that there is an even stronger dyadic effect when we isolate the influence of beliefs—meaning larger groups are particularly good at guessing the threshold when

bargaining with larger groups.

4. External Validity

A common concern with the use of laboratory experiments in political science has to do with the use of undergraduates as a convenience sample. The concern is that undergraduates are neither representative of "elite" decision-makers nor the general population from which they are drawn. As Renshon (2015) notes, such concerns are neither new nor unique to political science, as psychologists have long worried about the field's reliance on college students in drawing conclusions that may not be "externally valid". Renshon reviews a series of productive responses to these concerns, including attempts to replicate findings across different populations, with mixed results. In some studies, professionals/experts, behaved similarly to non-professionals/non-experts (Glaser et al. 2005), while in other cases the results were substantially different (Tyszka and Zielonka, 2002; Mintz et al. 2006). For example, Hafner-Burton et al. (2014) and LeVeck et al. (2014) found interesting differences between "elites" and student-subjects across a variety of strategic games, including the Ultimatum Game.

In many ways, we address this potential threat better than even the nascent experimental literature on crisis bargaining. In our experiment, we compare the behavior of individuals and groups in situations of ultimatum bargaining drawn from three different samples: undergraduate students at a top research university (allowing for comparison to previous work), subjects drawn from a more general population on Amazon Turk, and finally a sample of political "elites". We find important differences and surprising similarities across the different samples discussed above.

A second, and related concern is that subjects – both students and elites alike – would behave differently in real-life situations when compared to the lab. This could be because subjects are not fully motivated to engage in the experiment, or because factors in the real world that were omitted from the experiment may cause them to behave differently (similar to omitted variable bias when making inferences in observational studies). The latter is a constant risk with the use of experiments across all fields, including the biological sciences where scientists debate whether effects from "test tube" experiments conducted *in vitro,* are likely to generalize to highly complex living organisms *in vivo*. When studying decision-making processes, it is possible that important factors like experience, high-stakes, and emotions are relevant in the real world, even if not captured in the set-up of the experiment.

In the case of our experiment, there are at least two simplifications that may induce different results in the laboratory when compared to the real world. First, a reasonable case can be made that the voting mechanism in the experiment does not capture the intricacies of foreign policy decision-making in a democracy or any other state. This is true. Our voting rule – which calculates a group's proposal as the median proposal submitted in the group – purposely abstracts from factors like coalitional bargaining within states, democratic deliberation, and the influence of elite opinion leaders. This is not to assert that all decision-makers are completely independent in any real-world decision, but rather, to the extent individual inputs are at least somewhat independent, our treatment manipulation captures this independent component.

A second simplification is that bargaining scenario that subjects in our experiment face is

much simpler and lower stakes than the real-world bargaining scenarios faced by leaders.[22] This is also true. In closely matching our experiment to the assumptions of the bargaining model of war, we abstract from the multidimensional nature and high stakes of international crisis bargaining. However, there is compelling evidence that larger groups may still outperform individuals even if the situation were to become more complex[23] or if the stakes were raised in the domain of foreign policy. Indeed, whereas "average accuracy" in the model of collective accuracy outlined by Page et al. is likely to be low among the general population with respect to designing a rocket, in the domain of foreign policy, Tetlock (2005) has shown that – assuming non-specialists have some baseline knowledge of foreign affairs – "we reach the point of diminishing marginal predictive returns for knowledge disconcertingly quickly" in predicting what will happen in a particular region. That is, "average accuracy" of foreign affairs is typically at a sufficient level among "attentive readers of the New York *Times* in 'reading' emerging situations" to expect that even individual specialists are not significantly more reliable than groups of non-specialists. We expect that when the wisdom of the crowds is harnessed in the real world, that the larger, more diverse group of independently deciding individuals generally has some baseline accuracy. Moreover, laboratory evidence suggests that higher stakes have a fairly minimal effect on behavior in the Ultimatum Game (Camerer and Hogarth 1999).

---

[22] For example, in a multidimensional policy space, aggregating diverse preferences across multiple actors may result in a single foreign policy proposal that is ideologically incoherent – with more hawkish measures on some dimensions and more dovish measures on others.

[23] In fact, results from Hong and Page (2004) suggest the opposite. Groups of diverse individuals have a particular advantage in more complex decisions.

A third concern may be that it is much clearer what a "fair" or "acceptable" offer is in the Ultimatum Game – namely a 50-50 split – than it is in the real-world, where a "fair" or "acceptable" division can be much more ambiguous and contingent on factors that non-experts know little about (history, power, regime type, etc.).[24] If true, the structure of the Ultimatum Game may bias against the importance of expertise, by providing a clearer focal point around which the offers of non-expert proposers and responders can more easily converge when compared to the real world. This is certainly possible, and it is an interesting area for future research. However, we note that even in our relatively simple and controlled experiment, experienced individuals actually do perform better than inexperienced individuals, suggesting that the Ultimatum Game is not so simple that expertise is rendered meaningless. Instead, our results confirm that individual expertise helps, but they show that aggregation helps even more. This finding mimics related research showing that larger and more diverse groups of non-experts can outperform experts, even on complex issues related to foreign policy (Tetlock 2005). Furthermore, while the norm of 50-50 divisions is well known, there is good reason to suspect that it is not the only widely-known norm relevant for crisis bargaining. For example, work by Tomz and Weeks (2013) shows that citizens in different democratic states – the US and the UK – share many norms that are relevant to reducing the risk of conflict between democracies. It is possible that processes of aggregation could help distill which of these norms are most relevant to a particular crisis, and further reduce the chance of bargaining failure and war between democratic states.

Therefore, despite the fact that each of these three concerns are reasonable, we believe

[24] We thank an anonymous reviewer for raising this concern.

the level of realism in our experiment is appropriate for the specific hypotheses we seek to test. In general, we agree with McDermott that – rather than being a property of any individual experiment – "external validity follows, as replications across time and populations seek to delineate the extent to which...conclusions can generalize." (McDermott 2011, 28). Future studies can, and should, identify theoretically relevant conditions along which our experiment differs from the real world, and test – as part of a broader research program – whether the inclusion of these factors moderates the effects identified here.

5. Conclusion

The evidence we've gathered from our experiments are, of course, preliminary, and there is much more work that can be done in this area. This might include further studies that more systematically manipulate how information is distributed across individuals, the identity of bargainers, as well as the precise mechanism by which information is aggregated. Other studies may look at observational data to see how aggregated signals, such as market movements or polls, actually influence democratic decision-making. Finally, democracies and autocracies vary systematically in the caliber of various aggregation mechanisms (such as the depth of markets, or how informed their publics are). It is plausible that measures of this might be linked to measures of war-bargaining outcomes.

However, we believe that what we have been able to demonstrate so far is important. In bargaining scenarios that mimic key aspect of war-bargaining, aggregated offers from larger groups systematically outperform the offers made by smaller groups and individuals. Furthermore, part of the information aggregated appears to be individuals' knowledge of what they themselves would do in their opponent's shoes, and this may help them actually predict the

responses of their opponents. Thus, the democratic peace may partially arise because, similar to our study, democracies aggregate signals from diverse individuals, and those individuals provide more informative views when they are facing decision-makers who are similar to themselves.

These results notwithstanding, it is worthwhile to emphasize an important limitation to the inferences we seek to make from this study. To be clear, we do not claim that democracies always make better decisions in every situation. Indeed, there have been numerous cases where democratic decision-makers committed grave errors in crisis bargaining. For example, it is well documented that the United States made several errors in estimating the capabilities and resolve of Saddam Hussein in the run up to the Iraq War in 2003 (Gordon and Trainor 2006; Lake 2010). Examples like this suggest that – even if democracies can make collectively wiser decisions compared to non-democracies on average – they are certainly not immune from making decision errors in particular cases. However, we note that in many famous cases of miscalculation by democracies, the actual reason for the miscalculation is often attributed to more restricted decision-making, where a narrow group of similar-minded leaders engaged in an echo chamber (Janice 1972), and effectively excluded the diverse views of numerous individuals (Packer 2005, 50-60; Daalder and Lindsay, 46-47; Mann, 351-353).

References

Allison, Graham. 1969. Conceptual models and the Cuban missile crisis. *American Political Science Review*, *63*(03), 689-718.

Andreoni, James, and Emily Blanchard. 2006. Testing subgame perfection apart from fairness in ultimatum games. *Experimental Economics* 9(4): 307–321.

Bone, J., Hey, J. and Suckling, J. 1999. 'Are groups more (or less) consistent than individuals?', *Journal of Risk and Uncertainty*, vol. **8**, pp. 63–81.

Bornstein, Gary, and Ilan Yaniv. "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'Rational' Players?." *International Journal of Game Theory* 1(1): 101–8.

Braut-Hegghammer, Målfrid. 2016, Unclear Physics: Why Iraq and Libya Failed to Build Nuclear Weapons. Cornell University Press.

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36:309-41.

Bremer, Stuart A. 1993. "Democracy and Militarized Interstate Conflict, 1816-1965." *International Interactions* 18:231-49.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. "An Institutional Explanation for the Democratic Peace." *American Political Science Review* 93(4):791-807

Camerer, Colin. F., and Hogarth, Robin. M. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of risk and uncertainty*, *19*(1-3), 7-42.

Camerer, Colin. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cason, T. N. and Mui, V.-L. 1997. 'A laboratory study of group polarisation in the team dictator game', *Economic Journal*, vol. **107**, pp. 1465–83.

Chan, Steve. 1984. "Mirror, Mirror on the Wall ... are the Freer Countries More Pacific?" *Journal of Conflict Resolution* 28:617-48.

Daalder, Ivo H. and James M. Lindsay. 2003. America Unbound: The Bush Revolution in Foreign Policy. Washington, D.C.: Brookings Institution Press.

Dafoe, Allan. 2011. Statistical Critiques of the Democratic Peace: Caveat Emptor. *American Journal of Political Science*, *55*(2), 247-262.

Davis, J. H. 1992. 'Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: selected examples, 1950–1990', *Organizational Behaviour and Human Decision Processes*, vol. **52**, pp. 3–

38.DeNardo, James. *The amateur strategist: Intuitive deterrence theories and the politics of the nuclear arms race*. Cambridge University Press, 1995.

Dixon, William J. 1993. "Democracy and the Management of International Conflict." *Journal of Conflict Resolution* 37:42-68.

Dixon, William J. 1994. "Democracy and the Peaceful Settlement of International Conflict." *American Political Science Review* 88:1-17

Dobbs, Michael. 2008. *One Minute to Midnight*. Vintage.

Downs, Anthony. 1957. An Economic Theory of Democracy. New York: Columbia University Press.

Doyle, Michael W. 1997. *Ways of Ear and Peace: Realism, Liberalism, and Socialism* (p. 276). New York: Norton.

Elbittar, Alexander, Andrei Gomberg, and Laura Sour. 2011. "Group Decision-Making and Voting in Ultimatum Bargaining: an Experimental Study." *The B.E. Journal of Economic Analysis & Policy* 11(1).

Epley, Nicholas, Eugene Caruso, and Max H. Bazerman. "When perspective taking increases taking: reactive egoism in social interaction." *Journal of personality and social psychology* 91, no. 5 (2006): 872.

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

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379-414.

Fursenko, Aleksandr, and Timothy Naftali. 1998. " *One Hell of a Gamble": Khrushchev, Castro, and Kennedy, 1958-1964*. WW Norton & Company.

Fursenko, Aleksandr, and Timothy Naftali. 2007. *Khrushchev's Cold War: The Inside Story of an American Adversary*. WW Norton & Company.

Gartzke, Erik. 1998. "Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42 (January): 1-27.

Gartzke, Erik. 2000. Preferences and the Democratic Peace. *International Studies Quarterly*, *44*(2), 191-212.

Gartzke, Erik. 2007. The Capitalist Peace. *American Journal of Political Science*,*51*(1), 166-191.

Geddes, Barbara. 1999. Authoritarian Breakdown: Empirical Test of a Game-Theoretic Argument. Paper presented at the 95th Annual Meeting of the American Political Science Association, September, Atlanta

Geddes, Barbara. 2003. Paradigms and Sandcastles: Theory Building and Research Design in Comparative Politics. Ann Arbor: University of Michigan Press.

Gintis, Herbert. 2010. “Rationality and Common Knowledge.” Rationality and Society 22(3): 259–82.

Gleditsch, Nils Petter, and Håvard Hegre. 1997. "Peace and Democracy Three Levels of Analysis." *Journal of Conflict Resolution* 41, no. 2: 283-310.

Gleditsch, Kristian S, and Michael D. Ward. 1997. “Double Take a Reexamination of Democracy and Autocracy in Modern Polities.” *Journal of Conflict Resolution* 41(3): 361–83.

Gordon, Michael R. and Bernard E. Trainor. 2006 Cobra II: The Inside Story of the Invasion and Occupation of Iraq. New York: Pantheon.

Gorman, Martin J., and Alexander Krongard. 2005. *A Goldwater-Nichols act for the US government: Institutionalizing the interagency process*. Defense Intelligence Agency Washington Dc

Güth, Werner, Schmittberger, Rolf, and Schwarze, Bernd 1982. An Experimental Analysis of Ultimatum Bargaining. *Journal of economic behavior & organization*, *3*(4), 367-388.

Hafner-Burton, Emilie M, Brad L LeVeck, David G Victor, and James H. Fowler. 2014. “Decision Maker Preferences for International Legal Cooperation.” *International Organization* 68(04): 845–76.

Hafner-Burton, Emilie M., D. Alex Hughes, and David G. Victor. The Cognitive Revolution and the Political Psychology of Elite Decision Making. *Perspectives on Politics* 11, no. 02 (2013): 368-386.

Henisz, Witold J. 2000. “The Institutional Environment for Economic Growth.” *Economics & Politics* 12(1): 1–31.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. 2001. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review*, 73-78.

Hermann, Margaret G., and Charles F. Hermann. 1989. "Who makes foreign policy decisions and how: An empirical inquiry." *International Studies Quarterly* 33, no. 4: 361-387.

Hong, Lu, and Page, Scott E. 2004. Groups of Diverse Problem Solvers can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16385-16389.

Hong, Lu, and Scott Page. 2009. Interpreted and Generated Signals. *Journal of Economic Theory* 144 (5): 2174–2196.

Hong, Lu, and Scott Page. 2012. The Micro-Foundations of Collective Wisdom. In *Collective wisdom: principles and mechanisms*, edited by Hélène Landemore and Jon Elster. Collective Wisdom: Principles and Mechanisms.

Janis, Irving L. 1972. Victims of Groupthink: a Psychological Study of Foreign-Policy Decisions and Fiascoes. Oxford, England: Houghton Mifflin.

Maoz, Zeev, and Nasrin Abdolali. 1989. "Regime Types and International Conflict, 1816-1976." *Journal of Conflict Resolution* 33:3-36.

Kant, Immanuel. [1795] 1969. Perpetual Peace. Reprint New York: Columbia University Press

Kacowicz, Arie M. 1995. "Explaining Zones of Peace: Democracies as Satisfied Powers." *Journal of Peace Research* 32:265-76.

Kleindorfer, P. R., Kunreuther, H. C. and Schoemaker, P. J. H. 1993. Decision Science. An Integrative Perspective, New York: Cambridge University Press.

Kocher, Martin G., and Matthias Sutter. 2005. "The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. "*The Economic Journal* 115.500: 200-223.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of US Lawmaking*. University of Chicago Press.

Lake, David A. 1992. "Powerful Pacifists: Democratic States and War." *American Political Science Review* 86:24-37.

Lake, David A. 2011. Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War *International Security 3*5, no. 3: 7-52.

Lemke, Douglas, and William Reed. 1996. "Regime Types and Status Quo Evaluations: Power Transition Theory and the Democratic Peace." *International Interactions* 22:143-64

Lenz, Gabriel S. 2012. *Follow the Leader?* University of Chicago Press.

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: 18536-18541.

Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals became Democrats and Conservatives became Republicans*. University of Chicago Press.

Lupia, Arthur, and Mathew Mccubbins. 1994. “Learning From Oversight: Fire Alarms and Police Patrols Reconstructed.” *Journal of Law, Economics, and Organization* 10(1): 96.

Lupia, Arthur, and Matthew McCubbins. 2003. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge University Press.

Mann, James. 2004. Rise of the Vulcans: The History of Bush’s War Cabinet. New York: Penguin.

Maoz, Zeev, and Bruce Russett. 1992. "Alliance, Contiguity, Distance, Wealth, and Political Stability: Is the Lack of Conflict Among Democracies a Statistical Artifact?" *International Interactions* 17:245-68.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87:624-38.

Marcella, Gabriel. "National Security and the Interagency Process. 2004. " *US Army War College Guide to National Security Policy and Strategy* 239: 260.

Mullen, B., Anthony, T., Salas, E. and Driskell, J. E. 1994. 'Group cohesiveness and quality of decision making: an integration of tests of the Groupthink hypothesis', *Small Group Research*, vol. **25**, pp. 189–204.

Oneal, John R., and Bruce Russett. 1997a. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly* 41

Oneal, John R. and Bruce Russett. 1999. The Kantian peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885–1992.*World Politics*, *52*(01), 1-37.

Oneal, John R., and Bruce Russett. 2001. Clear and Clean: The Fixed Effects of the Liberal peace. *International Organization*, *55*(2), 469-485.

Packer, George. 2005. The Assassins' Gate: America in Iraq. New York: Farrar, Straus and Giroux

Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.

Popkin, Samuel L., 1991, *The Reasoning Voter,* Chicago, University of Chicago Press.

Powell, Robert. 2002. Bargaining Theory and International Conflict. *Annual Review of Political Science*, *5*(1), 1-30.

Raach, George T., and Ilana Kass. 1995. *National power and the interagency process*. National Defense Univ Washington Dc.

Rand, David G. 2012. "The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments." *Journal of Theoretical Biology* (0): 172–79.

Rand, David G., Tarnita, C. E., Ohtsuki, H., & Nowak, M. A. 2013. Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proceedings of the National Academy of Sciences*, *110*(7), 2581-2586.

Reiter, Dan, and Allan C. Stam. 2002. *Democracies at war*. Princeton University Press.

Renshon, Jonathan. 2015. Losing Face and Sinking Costs: Experimental Evidence on the Judgment of Political and Military Leaders. Forthcoming in *International Organization*

Rockenbach, B., Sadrieh, A. and Matauschek, B. 2001. 'Teams take the better risks', University of Erfurt, Working Paper.

Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97(4):585-602

Rousseau, David L., Christopher Gelpi, Dan Reiter, and Paul K. Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918-88." *American Political Science Review* 90:512-33.

Rummel, Rudolph J. 1983. "Libertarianism and International Violence." *Journal of Conflict Resolution* 27:27-71.

Rummel, Rudolph J. 1985. "Libertarian Propositions on Violence Within and Between Nations: A Test Against Published Research Results." *Journal of Conflict Resolution* 29:419-55.

Russett, Bruce. 1993. Grasping the Democratic Peace: Principles for a Post-Cold War World. Princeton, NJ: Princeton University Press.

Russett, Bruce., Oneal, John R., & Davis, David R. 1998. The third leg of the Kantian tripod for peace: International organizations and militarized disputes, 1950–85.*International Organization*, *52*(3), 441-467.

Saunders, Elizabeth Nathan. *Leaders at war: how presidents shape military interventions*. Cornell University Press, 2011.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 94(4):829-44.

Schultz, Kenneth A. 2001. Democracy and Coercive Diplomacy.Cambridge: Cambridge University Press.

Signorino, Curtis. S. 1999. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review*, *93*, 279-298.

Slantchev, Branislav L. "Feigning weakness." *International Organization* 64, no. 03 (2010): 357-388.

Small, Mel., and J. David Singer. 1976. "The War Proneness of Democratic Regimes, 1816-1965." *The Jerusalem Journal of International Relations* 1:50-69.

Sniezek, J. A. 1992. 'Groups under uncertainty: an examination of confidence in group decision making', *Organizational Behaviour and Human Decision Processes*, vol. **52**, pp. 124–55.

Stevenson, Hayley. 2016, "The Wisdom of the Many in Global Governance: An Epistemic-Democratic Defense of Diversity and Inclusion." *International Studies Quarterly* 60.3: 400-412.

Surowiecki, James. 2005. *The wisdom of crowds*. Random House LLC.

Taubman, William. 2003. *Khrushchev: the man and his era*. WW Norton & Company.

Tetlock, Philip. 2005. *Expert political judgment: How good is it? How can we know?*. Princeton University Press.

Thompson, William R., and Richard Tucker. 1997. "A Tale of Two Democratic Peace Critiques." *Journal of Conflict Resolution* 41:

Tingley, Dustin H, and Stephanie W. Wang. 2010. “Belief Updating in Sequential Games of Two-Sided Incomplete Information: an Experimental Study of a Crisis Bargaining Model.” *Quarterly Journal of Political Science* 5(3): 243–55.

Tingley, Dustin, and Barbara Walter. 2011. “Can Cheap Talk Deter? an Experimental Analysis.” *Journal of Conflict Resolution* 55(6).

Tomz, Michael. 2007. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization*, *61*(4), 821.

Tomz, Michael R., and Jessica LP Weeks. 2013. "Public opinion and the democratic peace." *American Political Science Review* 107, no. 04: 849-865.

Tsebelis, George, and Seung-Whan Choi. 2009. “Democratic Peace Revisited: It Is Veto Players.” *Presented at the American Political Science Association Conference in Toronto*.

Weede, Erich. 1984. "Democracy and War Involvement." *Journal of Conflict Resolution* 28:649-64.

Weede, Erich. 1992. "Some Simple Calculations on Democracy and War Involvement." *Journal of Peace Research* 29:377-83.

Weeks, Jessica L. 2008. Autocratic audience costs: Regime Type and Signaling Resolve. *International Organization*, 35-64.

Weeks, Jessica L. 2012. Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict. American Political Science Review. 106.2

Weeks, Jessica L. 2014. Dictators at War and Peace. Cornell Studies in Security Affairs, Cornell University Press.

Wolfers, Justin, and Zitzewitz, Eric. 2009. Using Markets to Inform Policy: The case of the Iraq war. *Economica*, *76*(302), 225-250.

Young, H. Peyton. 1993. “The Evolution of Conventions.” Econometrica 61(1): 57–84.

Young, H. Peyton, and Mary A. Burke. 2001. “Competition and Custom in Economic Contracts: a Case Study of Illinois Agriculture.” The American Economic Review 91(3): 559–73.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion (Cambridge Studies in Public Opinion and Political Psychology)*. Cambridge University Press.