Supplementary Information

Democratic Peace and the Wisdom of Crowds

Experimental Instructions 1
Recruitment of Subjects and Analysis of Balance 8
Side B Earnings 11
Analysis of Group Advantage at Different Magnitudes 12
Why our Results are Consistent with a Dyadic Effect 13
Quantitative Differences in Number of Actors Influencing Foreign Policy Across Autocracies and Democracies 16
Removing the Normative Component to Group Advantages in Formulating Accurate Beliefs about Minimum Acceptable Offers 17
Extended Literature Review of When Groups are *Not* Wiser 19
Supplementary Information References 21

Experimental Instructions

The main experiment used in our paper was conducted in March of 2015. Subjects were recruited and paid through the online labor market Amazon Mechanical Turk, and took the study through Qualtrics, a web based program for administering randomized survey questions. The instructions for the main experiment are below. Prior, to completing this portion of the survey, subjects had the option of voluntarily answering a small set of demographic question about their gender, age, and education. After completing the main portion of our experiment, subjects answered a short comprehension quiz, which is also listed below.

The main text of our paper also reports results from a set of undergraduate subjects and international elites. These samples were originally collected by LeVeck, Hughes, Fowler Hafner-Burton, and Victor (2014). Instructions used for these samples can be found in that article's Supplementary Information Appendix.

Text of Main Experiment

All portions in square brackets were randomized. All numerical examples were drawn from an appropriate uniform distribution, so as not to systematically bias subjects' responses.

Screen 1

For this task, you have been randomly assigned by a computer to be a member of Side [A/B].
Other participants in this task have been assigned to be a member of Side [B/A].

Side A has [1/3/9] members.
Side B has [1/3/9] members.

Task: In this task *each member* of *each side* has an opportunity to earn a bonus between $0.00 and $0.40.

- **Members of Side A** will vote on how much of $0.40 should be offered to *each* member of Side B, with the remainder of $0.40 going to *each* member of Side A (see examples below).
- **Members of Side B** will vote on whether to accept or reject Side A's offer (see examples below).

Here is how voting works for each side:

Voting for Members of Side A

- Each member of Side A votes on every possible amount that could be offered to Side B (in 2-cent increments).
- Side A's actual offer to Side B is the largest offer supported by a majority of Side A's members. A majority of Side A is at least [1/2/5] out of [1/3/9] members.
    - **For example**, if the largest offer supported by a majority of Side A is $[0.01], then Side A proposes that *each* member of Side B receive $[0.01] and that *each* member of Side A receive the remainder ($[0.39]).

Voting for Members of Side B

- Each member of Side B votes on whether to accept or reject on every possible offer that Side A could make.
- If a majority of Side B's members vote to accept the offer that Side A actually makes, then the offer is accepted and goes through as proposed. A majority of Side B is at least [1/2/5] out of [1/3/9] members.
    - **For example** if Side A proposes that each member of Side B receive $[0.01] and a majority of Side B votes to accept this proposal, then *each* member of Side B receive $[0.01] and *each* member of Side A receives the remainder, ($[0.39]).
- If a majority of Side B's members vote to reject the proposal that Side A actually makes, then all players on each side earn $0.00.
    - **For example** if Side A proposes that each member of Side B receive $[0.01], and a majority of Side B votes to reject this proposal, then *each* member of Side B receives $0.00 and *each* member of Side A receives $0.00.

Screen 2 (Side A)

Ok, time to vote on what Side A should offer to Side B.

- Side A's offer will be the *largest* offer to Side B that is supported by a majority of Side A's 1 members.
    - If a majority of Side B's [1/3/9] member[s] vote to accept this offer, it will be implemented.
    - If a majority of Side B's [1/3/9] member[s] vote to reject the offer, each member of both sides receives $0.00

As a member of Side A, please vote by selecting the largest offer you support making to Side B. By picking an offer, you are voting for that specific offer and for every lesser offer to Side B. Side A's final offer will be the largest offer to Side B that is supported by a majority of Side A's members.

- ○ Side A offers that each member of Side B receive **$0.00**, meaning that each member of Side A receives $0.40
- ○ Side A offers that each member of Side B receive **$0.02**, meaning that each member of Side A receives $0.38
- ○ Side A offers that each member of Side B receive **$0.04**, meaning that each member of Side A receives $0.36
- ○ Side A offers that each member of Side B receive **$0.06**, meaning that each member of Side A receives $0.34
- ○ Side A offers that each member of Side B receive **$0.08**, meaning that each member of Side A receives $0.32
- ○ Side A offers that each member of Side B receive **$0.10**, meaning that each member of Side A receives $0.30
- ○ Side A offers that each member of Side B receive **$0.12**, meaning that each member of Side A receives $0.28
- ○ Side A offers that each member of Side B receive **$0.14**, meaning that each member of Side A receives $0.26
- ○ Side A offers that each member of Side B receive **$0.16**, meaning that each member of Side A receives $0.24
- ○ Side A offers that each member of Side B receive **$0.18**, meaning that each member of Side A receives $0.22
- ○ Side A offers that each member of Side B receive **$0.20**, meaning that each member of Side A receives $0.20
- ○ Side A offers that each member of Side B receive **$0.22**, meaning that each member of Side A receives $0.18
- ○ Side A offers that each member of Side B receive **$0.24**, meaning that each member of Side A receives $0.16
- ○ Side A offers that each member of Side B receive **$0.26**, meaning that each member of Side A receives $0.14
- ○ Side A offers that each member of Side B receive **$0.28**, meaning that each member of Side A receives $0.12
- ○ Side A offers that each member of Side B receive **$0.30**, meaning that each member of Side A receives $0.10
- ○ Side A offers that each member of Side B receive **$0.32**, meaning that each member of Side A receives $0.08
- ○ Side A offers that each member of Side B receive **$0.34**, meaning that each member of Side A receives $0.06
- ○ Side A offers that each member of Side B receive **$0.36**, meaning that each member of Side A receives $0.04

- ○ Side A offers that each member of Side B receive **$0.38**, meaning that each member of Side A receives $0.02
- ○ Side A offers that each member of Side B receive **$0.40**, meaning that each member of Side A receives $0.00

Screen 2 (Side B)

Ok, time to vote on whether to accept or reject Side A's offer.

- Side A's offer will be the *largest* offer to Side B that is supported by a majority of Side A's 1 members**.**
    - If a majority of Side B's [1/3/9] members vote to accept this offer, it will be implemented.
    - If a majority of Side B's [1/3/9] members vote to reject the offer, each member of both sides receives $0.00

	Vote to Accept this Offer	Vote to Reject this Offer
Side A offers that each member of Side B receive **$0.00**, meaning that each member of Side A receives $0.40	❍	❍
Side A offers that each member of Side B receive **$0.02**, meaning that each member of Side A receives $0.38	❍	❍
Side A offers that each member of Side B receive **$0.04**, meaning that each member of Side A receives $0.36	❍	❍
Side A offers that each member of Side B receive **$0.06**, meaning that each member of Side A receives $0.34	❍	❍
Side A offers that each member of Side B receive **$0.08**, meaning that each member of Side A receives $0.32	❍	❍
Side A offers that each member of Side B receive **$0.10**, meaning that each member of Side A receives $0.30	❍	❍
Side A offers that each member of Side B receive **$0.12**, meaning that each member of Side A receives $0.28	❍	❍
Side A offers that each member of Side B receive **$0.14**, meaning that each member of Side A receives $0.26	❍	❍
Side A offers that each member of Side B receive **$0.16**, meaning that each member of Side A receives $0.24	❍	❍
Side A offers that each member of Side B receive **$0.18**, meaning that each member of Side A receives $0.22	❍	❍
Side A offers that each member of Side B receive **$0.20**, meaning that each member of Side A receives $0.20	❍	❍
Side A offers that each member of Side B receive **$0.22**, meaning that each member of Side A receives $0.18	❍	❍
Side A offers that each member of Side B receive **$0.24**, meaning that each member of Side A receives $0.16	❍	❍
Side A offers that each member of Side B receive **$0.26**, meaning that each member of Side A receives $0.14	❍	❍
Side A offers that each member of Side B receive **$0.28**, meaning that each member of Side A receives $0.12	❍	❍
Side A offers that each member of Side B receive **$0.30**, meaning that each member of Side A receives $0.10	❍	❍
Side A offers that each member of Side B receive **$0.32**, meaning that each member of Side A receives $0.08	❍	❍
Side A offers that each member of Side B receive **$0.34**,	❍	❍

meaning that each member of Side A receives $0.06		
Side A offers that each member of Side B receive **$0.36**, meaning that each member of Side A receives $0.04	❍	❍
Side A offers that each member of Side B receive **$0.38**, meaning that each member of Side A receives $0.02	❍	❍
Side A offers that each member of Side B receive **$0.40**, meaning that each member of Side A receives $0.00	❍	❍

Comprehension Quiz

Which side are you a member of?

- ❍ Side A
- ❍ Side B

How many individuals are members of Side A? Use the slider below to enter your response.

How many individuals are members of Side B? Use the slider below to enter your response.

If Side A offers 10 cents (out of 40 cents) to each member Side B and a majority of Side B vote to accept this offer:

What bonus does each member of Side A receive (in cents)?
What bonus does each member of Side B receive (in cents)?

If Side A offers 10 cents (out of 40 cents) to each member Side B and a majority of Side B vote to reject this offer:

What bonus does each member of Side A receive (in cents)?
What bonus does each member of Side B receive (in cents)?

Recruitment of Subjects and Analysis of Balance

For our main experiment, we initially recruited 2326 subjects from the online labor market Amazon Mechanical Turk (AMT). As stated in the main text, subjects were paid a $0.50 show-up fee, plus a bonus depending on their performance in the game. Upon arriving at our study's webpage, subjects were randomly assigned to one of the 5 conditions: 1-1, 3-3, 3-9, 9-3, 9-9, where the first number denotes the number of proposers and the second number denotes the number of responders on a given side. Subjects were then randomly assigned to a side, A or B, for a total of 10 cells: A1-1 A3-3 A3-9 A9-3 A9-9 B1-1 B3-3 B3-9 B9-3 B9-9.

In the main paper, we estimate average performance by repeatedly sampling groups in each condition with replacement, and evaluating their average performance. This is valid since each subject's decision was made independently of all other subjects in a particular condition, and each player in a condition made a choice in light of the same incentives and information. However, because subjects were told they were randomly matched with a particular group, we wanted to pay them accordingly. We therefore had to create and match a fixed set of actual groups prior to paying subjects. In order to allow players to be matched in this manner, we randomly allocated 3x as many proposers in the A9-3 condition and 3x as many responders in the B3-9 condition. There were a few subjects in each condition that could not be matched for payment because randomization did not create perfect proportions necessary to form and match unique groups. These subjects were debriefed through AMT's

messaging interface that we had not matched them, and were paid their full show-up fee, plus their full possible bonus. However, at the time these subjects made their choice, they did not know they could not be matched. We therefore use their choices in our analyses.

To be included in the sample analyzed in the main paper, subjects also had to pass a comprehension quiz (above), which helped to ensure that they had read and understood the instructions. Of the people who participated 1409 subjects passed this quiz, making the passage rate 60.6%, which is very close to another study on the ultimatum game which used an AMT sample and included a quiz that was very similar to the quiz employed in our study (Rand et al. 2013).

However, our analyses show that passage of the quiz was not quite random. In particular, a smaller number of subjects passed the quiz in the 3-9 and 9-3 conditions because these subjects were more likely to misremember the number of subjects on each side. It seems unlikely that this fact would drive the results in our paper, but we can check this because subjects took the quiz after voting.

First, we can show that including subjects who incorrectly stated the number of players on one or more sides restores balance:

Appendix Table A1: Distribution of subjects who passed the quiz vs. theoretical distribution (second row)

A1-1	A3-3	A3-9	A9-3	A9-9	B1-1	B3-3	B3-9	B9-3	B9-9
123	124	85	286	92	109	110	280	98	102
101	101	101	302	101	101	101	302	101	101

Compared to the theoretical randomization distribution on the second row of Table 1[1], the distribution of subjects who passed the quiz is significantly different (χ_2= 17.65, df = 9, p-value = 0.04)

However, Table 2, below, shows the distribution of subjects after we include those who misremembered the number subjects on at least one side, and this is not significantly different from the theoretical randomization distribution, shown on the second row[2] (χ_2= 11.77, df = 9, p-value = 0.23).

[1] Rounded for display purposes.
[2] Rounded for display purposes.

Appendix Table A2: Distribution of subjects who passed the quiz or incorrectly stated the number of players on at least one side vs. theoretical distribution (second row)

A1-1	A3-3	A3-9	A9-3	A9-9	B1-1	B3-3	B3-9	B9-3	B9-9
143	139	107	368	105	136	118	372	124	118
124	124	124	371	124	124	124	371	124	124

We can then proceed to re-run the analyses in the main text on this modified sample and compare them to the results in our main text. Table 3 compares the proportion of bargaining successes across conditions, while Table 4 compares Side A's average earnings across each condition. Success rates and earnings are slightly lower in the sample that includes people who answered the quiz incorrectly, and the standard errors are slightly higher in this sample. Both are consistent with what one would expect from a pool of subjects who hold less accurate views about the game they are playing. However, the reduced rate of success and earnings is small and fairly constant across the board, and the pattern of responses remains the same as the pattern reported in the main text.

Appendix Table A3: Proportion of bargaining successes in subjects who passed the comprehension quiz, and subjects who incorrectly stated the number of players

Condition	Passed the quiz (Presented in the manuscript)	Incorrectly stated the # of players
3-3	0.76 (0.050)	0.74 (0.050)
3-9	0.87 (0.042)	0.87 (0.042)
9-3	0.90 (0.029)	0.88 (0.031)
9-9	0.98 (0.025)	0.95 (0.039)

Bootstrapped standard error below in parentheses

Appendix Table A4: Average earnings (in cents) for Side A among subjects who passed the comprehension quiz, and subjects who incorrectly stated the number of players

Condition	Passed the quiz (Presented in the manuscript)	Incorrectly stated the # of players
33	15.87	15.42
	(0.965)	(.960)
39	17.85	17.04
	(0.780)	(0.783)
93	18.71	18.13
	(0.522)	(0.565)
99	19.44	19.16
	(0.479)	(0.479)

Bootstrapped standard error of the mean below in parentheses

One can further examine whether the sample that passed the quiz shows severe imbalance along any of the 3 pretreatment quiz covariates we collected. Prior to the experiment, we asked subjects to self report their gender (male, female), level of education (Grammar, HS/GED, Vocational/Technical, Some College, 4 yr College, Masters, Ph.D, Professional), Age (under 13, 13-17, 18-25, 26-34, 35-54, 55-64, 65+). None of the p-values from a Kruskal-Wallis one-way analysis of variance test across conditions were significant at conventional levels (Gender $\chi_2 = 14.57$ df = 9, p = 0.103; Age $\chi_2 = 5.06$ df = 54, p = 0.823; Education $\chi_2 = 2.56$ df = 72, p = 0.978).

Side B Earnings

In the main text of the paper, we focus on the number of times Side A makes a successful proposal in each of our main conditions 33, 39, 93 and 99, where the first number represents the number of individuals on Side A and the second number represents the number of individuals on Side B. This does not present a problem for measuring bargaining successes, since success happens for both sides in the Ultimatum Game. However, it is not guaranteed that just because Side A earns more in one condition that Side B will do the same. In Table 5 below, we present evidence that this is the case. As the number of players on one or more sides grows, Side B players earn more.

Appendix Table A5: Average earnings (in cents) for Side A and Side B

Condition	Side A earnings	Side B Earnings
33	15.87 (0.965)	14.70 (1.069)
39	17.85 (0.780)	17.48 (0.907)
93	18.71 (0.522)	17.74 (0.643)
99	19.44 (0.479)	19.32 (0.605)

Bootstrapped standard error of the mean below in parentheses

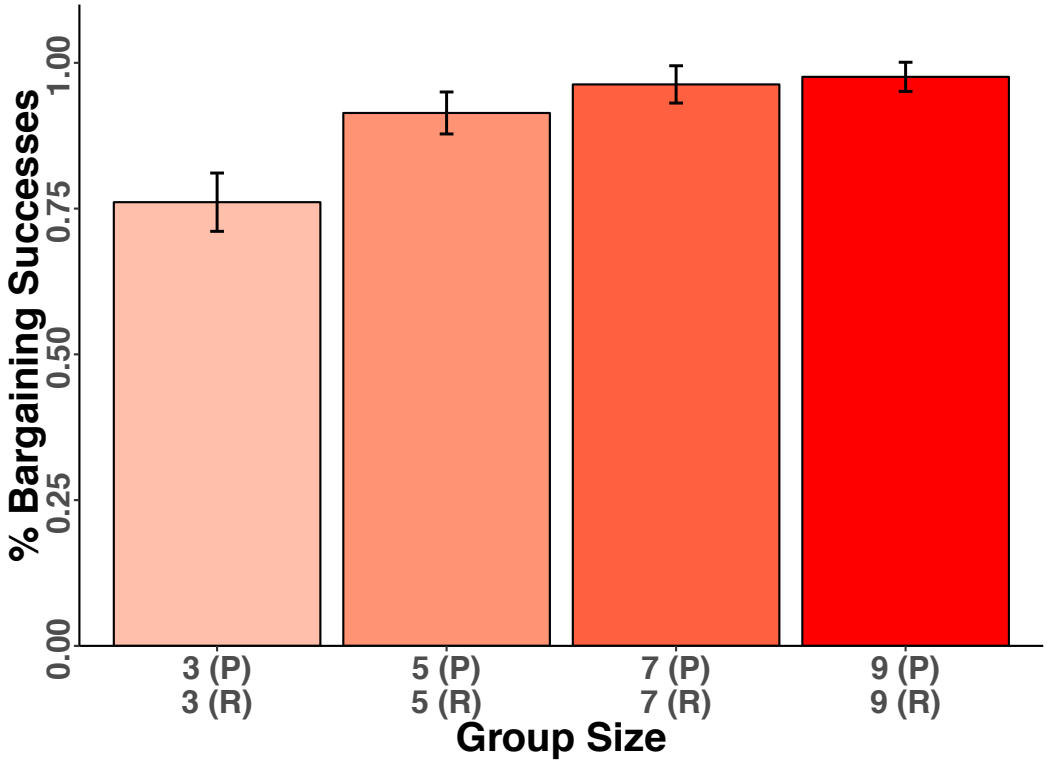
Analysis of Group Advantage at Different Magnitudes

In our experiment, we test a simple directional hypothesis that larger groups of independent decision-makers produce better bargaining outcomes. Our results confirm this hypothesis, especially when comparing democratic dyads (9 proposers vs. 9 responders) to autocratic dyads (3 proposers vs. 3 responders). The relative size of the groups representing autocracies and democracies were chosen largely for practical reasons, and there is, to our knowledge, no literature stating that most democracies are 3x larger than autocracies in terms of how many decision-makers independently influence pre war bargaining. Instead this relatively large group-size differential was chosen to be 3x because we wanted to ensure that we would detect an effect if our directional hypothesis was, in fact, correct (and reject it if it were not). That is, we wanted to ensure that a null result would not simply be the result of using too small of a "dosage".

However, it is useful to examine how sensitive our result is to a particular group size. Would our results have differed substantially if we had defined democratic dyads as being groups of 5 or 7 rather than 9? While it would be too costly to re-run the experiment with every possible permutation, we used our data to simulate what would happen at these smaller group sizes. Specifically we took subjects from the 9 vs. 9 democratic dyad condition and used these subjects to form groups of 5 vs. 5 and 7 vs. 7. The key assumption behind this simulation is that subjects would not act very differently if they knew they were in groups smaller than 9 and larger than 3.

Figure 1 below shows the results with vertical lines representing the standard error of the mean. We see a dramatic increase in bargaining success going from groups of 3 (autocracies) to groups of 5. Unlike autocracies, which avoid bargaining failure 76% of the time, dyads of 5 vs 5 avoid it 91% (95% c.i. 84% to 98%, *t*-test difference in means $p < .001$). Therefore, our finding does not seem particularly sensitive to the particular size ratio used in the experiment to distinguish democracies from autocracies (3x). A similar result would have obtained had we used a ratio as low as 1.66x.

Appendix Figure A1: Simulations of Bargaining Success for Group sizes between 3 and 9.



Why our Results are Consistent with a Dyadic Effect

In the main body of the text, we note that our main finding appears to be monotonic, such that autocratic dyads do the worst (no better than individuals), mixed dyads may do better, and democratic dyads do the best. Taken by itself, this monotonic effect appears more consistent with a monadic peace, whereby democracies always do better than a dyadic peace (where democracies do substantially better only when interacting with other democracies).

However, we note that there at least two ways that our findings could be consistent with the dyadic peace finding in the observational literature. First, mixed dyads may have an overall

higher rate of dispute initiation that offsets the benefit of aggregation. Second, factors not present in our experiment could lead members of mixed dyads to have systematically biased views about how to bargain with one another type, and this could cause aggregation to actually produce worse bargaining outcomes in mixed dyads. Here, we elaborate on each of these possibilities.

(1) Differences in Dispute Initiation

It is possible that mixed dyads engage in crisis bargaining more often due to more frequent conflicts of interest. Thus, if mixed dyads are indeed more war-prone than even autocratic dyads, our results suggest that this may be due to a higher absolute number of crisis bargaining *opportunities* between autocratic states compared to mixed dyads (Lake 1992), which would lead to a higher rate of war participation despite a lower average rate of bargaining failure per interaction

Meanwhile, our experiment assumes away differences in crisis/dispute initiation in order to compare the relative effectiveness of larger groups to smaller groups *within* bargaining situations. That is, we force all participants into crisis bargaining. If, as the literature has found, there are differences in the rates at which mixed and non-mixed dyads initiate crises/disputes (and hence pre-war bargaining) (Lake 1992), then our results are not inconsistent with a dyadic peace. This is because the literature has found that mixed dyads initiate more disputes than non-mixed dyads.

Due to space constraints in the main manuscript, we did not give a concrete example to illustrate our argument. Here we do so with a simple numerical example:

Assume that, conditional on a dispute, democracies have a 10 percent lower chance of bargaining failure, and this reduction is, at least partially, additive, such that 2 democracies have between an 11% and 20% lower chance of going to war over a dispute (compared to an autocratic dyad). Because the effect is additive, you will expect to see a monotonic effect—where mixed dyads go to war 10% less than autocratic dyads, but more often than democratic dyads. This seems contrary to the dyadic version of the democratic peace where mixed dyads engage in conflict at roughly the same rate compared to autocratic dyads. Only democratic dyads have substantially rates of war.

However, this monotonic effect only exists in the example above because we are already conditioning on the existence of a dispute. Meanwhile the empirical literature says that mixed dyads initiate many more disputes compared to democratic dyads or autocratic dyads. Adding in this factor can generate a dyadic version of the democratic peace. For example, what if the probability of an autocratic dyad going to war is 10%, conditional on a dispute, and the probability of a dispute between two autocracies is 10% in a given year, making the overall probability 1%. Further assume that the dispute rate is the same for democratic dyads, but the probability of war conditional on a dispute is 20% less. So democratic dyads go to war 0.8% of the time. If mixed dyads had the same dispute rate, but a 10% lower rate of war conditional on the dispute, they would go to war 0.9% of the time. However, if mixed dyads have at least an 11% higher dispute rate (compared to non-mixed dyads), the advantage is effectively wiped out, and mixed dyads will have at least the same overall prevalence of war

as autocratic dyads. Therefore, a dyadic peace emerges in this case because there is a multiplicative effect between the fact that democracies are better at avoiding bargaining failure and non-mixed dyads are better at avoiding pre-war bargaining in the first place.

(2) Inaccurate Views about Different Regimes/Societies

Our experiment was conducted using decision-makers who all come from the same society (the US). Meanwhile, members of mixed dyads come from different societies which often have different governing structures, values, and norms. Previous studies have shown that such differences can lead to less accurate views about how to bargain with one another (Young and Burke 2001), and this can lead to higher rates of bargaining failure in ultimatum games between individual decision-makers (Chuah et al. 2007).

If members of mixed dyads are systematically worse at bargaining with one another, then information aggregation in the real world may not have the monotonic effect observed in our experiment. Instead, aggregating a larger number of systematically inaccurate views could make decision-making worse (compared to individual decision-making) (Austen-Smith and Banks 1996; Page 2008). In this case, the Wisdom of Crowds would, by itself, produce a strictly dyadic peace, as aggregation would help democracies bargain with one another, but might make democracies worse at bargain with autocracies (compared to autocratic dyads).

While our experiment does not directly test this explanation, some of our findings are, in fact, consistent with this proposition. Specifically, in Figure A2 below, we show that subjects appear to arrive at their offers by introspecting about what they themselves would accept. Such introspection is common in strategic decision-making (Coricelli and Nagel 2009; Epley, Caruso, and Bazerman 2006; Krueger and Acevedo 2008), but it might lead people astray when they try to put themselves in the shoes of people who are very different from themselves, either because decision-makers incorrectly impute their own preferences onto others, or because they filter their introspection through incorrect stereotypes about members of a different group (Epley et al. 2004; Epley, Caruso, and Bazerman 2006). Recent research suggests that subjects are not able to fully remove these biases, even when they have a clear incentive to do so and even when they are able to communicate with one another (Epley, Caruso, and Bazerman 2006; Epley, Keysar, Boven, and Gilovich 2004).

This explanation could be directly tested in future research by having individuals from countries with different regime types bargain with one another, while manipulating group size (as in our experiment).

Quantitative Differences in Number of Actors Influencing Foreign Policy Across Autocracies and Democracies

In the main manuscript we note that autocracies and democracies differ in the number of institutional actors who might influence policy-making—including foreign policy. However, we did not have space to display this comparison, so we do so here.

In Table 6 below, we show how Democracies (defined as 6 or higher on the Polity IV scale) have higher levels on both the *XCONST* Variable and the POLCONIII index.

Appendix Table A6: Difference in Institutional Constraints and Veto Points Between Democracies and Autocracies

Regime	Mean XCONST	Mean POLCONIII
Democracy	6.31	0.40
Autocracy	0.76	0.01

The Polity IV index, on which much of the democratic peace finding is based, is mostly driven by the variable *XCONST*, which codes the level of institutional constraints on the executive (Gleditsch and Ward 1997).This variable partially reflects the larger number of actors that constrain the executive when it comes to policy-making. XCONST therefore serves as an imperfect proxy for the fact that democratic policy making is typically influenced by a larger number of independent actors.

Meanwhile, the Polity IV index is also highly correlated with the POLCONIII index (r = .79, *Spearman's rho* = .77), which measures on a -1 to +1 scale the effective number of veto players in a polity (Henisz 2000) by using the raw number of institutional veto players and their relative independence in terms of preferences and ideological view points. Of all the available cross-country measures, this index most closely captures our independent variable, the number of independent decision-makers with inputs into the policy making process. [3]

[3] Interestingly, recent research also finds that while POLCONIII is highly correlated with Polity IV, it may do more to explain the prevalence of intrastate conflict than the Polity IV index (Tsebelis and Choi 2009). This is certainly consistent with our argument that the Democratic peace finding may be driven by the larger number of relevant decision-makers found in democracies.

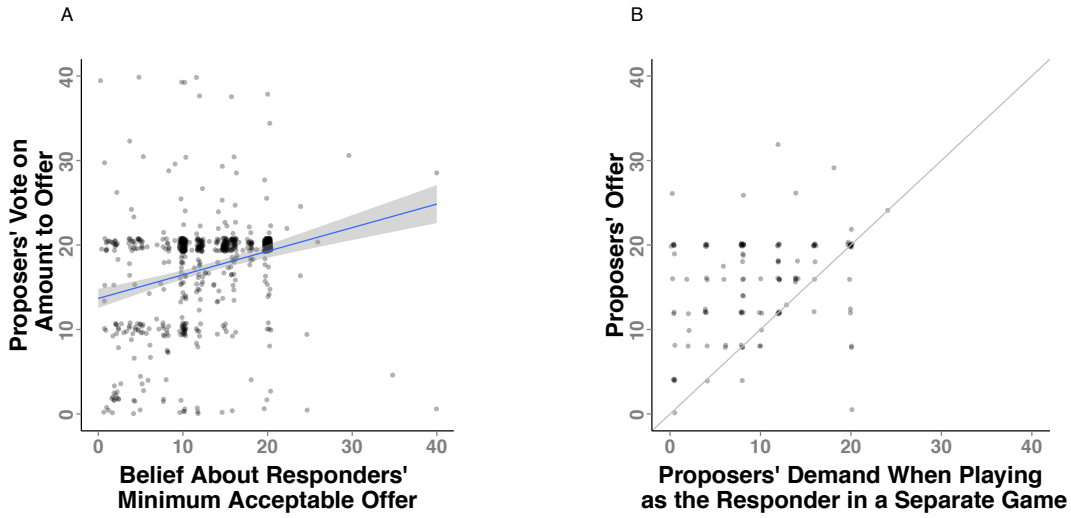
While prior research has shown that executives are often less constrained by other governmental actors in the domain of foreign affairs (Canes-Wrone, Howell, and Lewis 2008), it also shows that other veto players in government (such as the legislative branch) still have some influence over the executive's actions in the lead up to war and other foreign interventions (Canes-Wrone, Howell, and Lewis 2008; Howell and Pevehouse 2005; 2007; 2011). As a result, the greater presence of these veto players in democracies compared to autocracies should still increase the relative number of decision-making inputs—even if the influence of these additional actors is not as great in foreign policy as it is in domestic policy. In the main manuscript, we also review how a greater number of effective actors may exist within the staff of democratic executives themselves.

Removing the Normative Component to Group Advantages in Formulating Accurate Beliefs about Minimum Acceptable Offers

A third aspect we investigated, but could not include in the main manuscript is what factors are actually driving the observed behavior in our experiment. Part of what could be driving the effect could be, as we have proposed, beliefs about what the other side should accept, but part of the effect could also be subjects simply acting according to a particular norm. For instance a modal offer in many games is to offer a 50-50 split, which may reflect a societal norm about what is fair (Camerer 2003).

To measure Proposers' beliefs about what Responders would accept, we asked subjects to guess what the other side's minimum acceptable offer would be in a post experiment survey. To give subjects some incentive for accuracy, we paid subjects an additional 0.25 USD if they were correct, 0.05 USD if they were within 2 cents of the correct guess, and 0.01 USD if they were within 4 cents of the correct guess. The relationship between these guesses and proposers' actual votes in the experiment is show in Panel A of **Figure 3**. Plot A in this panel shows that subjects' votes are, in fact, related to their beliefs about what the other side will accept, though there is also a visible tendency to offer a 50/50 split even when subjects believe that the other side would accept a lower offer. For example, many subjects thought the other side would accept as little as 10 cents, but still offered half of the pie (20 cents).

Appendix Figure A2: Beliefs versus Offers



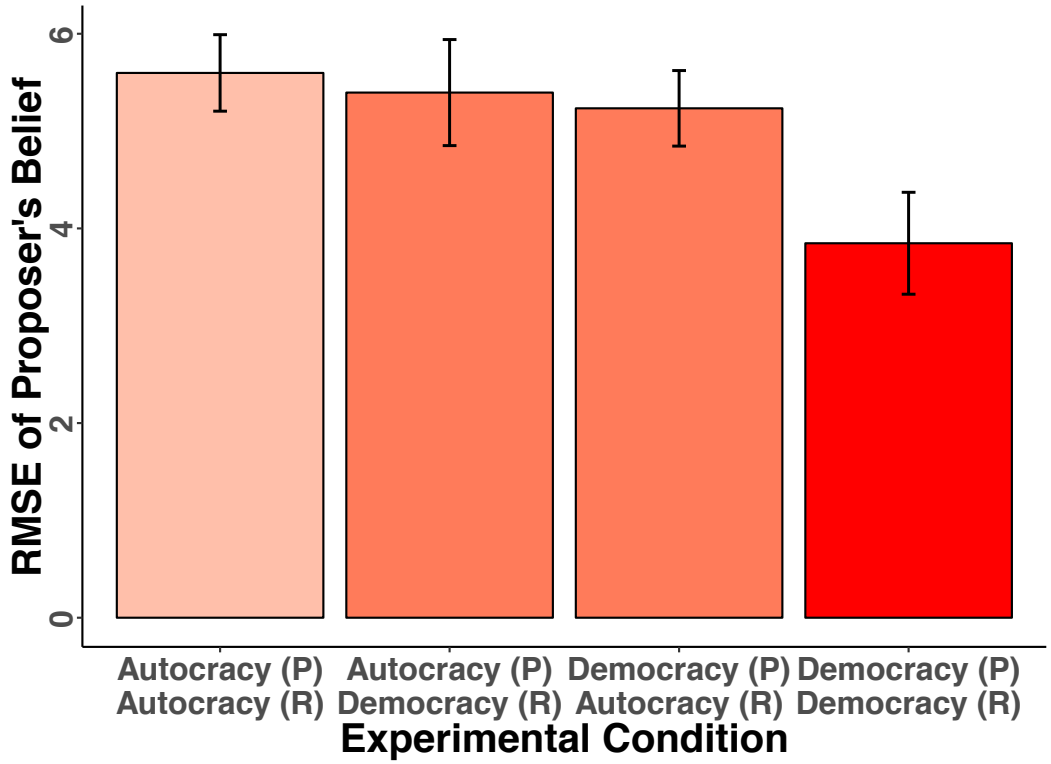
Given this finding, it is reasonable to suspect that subjects' votes might be a combination of introspection (thinking about what they would accept if they were in the responder's shoes) and norms of fairness (wanting to make a 50/50 split). To examine this we looked at data from a separate experiment where 132 subjects had played the ultimatum game twice (LeVeck et al. 2014), both as a proposer and as a responder. **Panel B** of **Figure 3** shows the relationship between these players' proposals (plotted on the y axis) and their demands (plotted on the x axis). The data shows that the vast majority of subjects' offers are between a 50/50 division and what they would demand in the same game. This is suggestive that, in a population where individuals likely consider themselves to be typical, they take their own type to provide information about the distribution of types they are facing. That is, they put themselves in the other player's shoes and ask, "What would I demand if I were them?"

Our aggregation mechanism therefore seems to effectively aggregate two distinct factors: normative behaviors and belief about what the other side's minimum acceptable offer will be. Because our theory is focused around the second element – knowledge/information about the threshold for bargaining failure – we isolated that component of individual decisions to see if our main hypothesis would hold if people had simply acted on this knowledge (rather than on norms as well). In Figure 3, we analyze how accurate subjects were at guessing this threshold in terms of the root mean squared error of their guess. The results of this analysis show that there is an even stronger dyadic effect, where larger groups are particularly good at guessing the threshold when bargaining with larger groups. We also find that that larger groups (especially democratic dyads) are much better at actually guessing the responders' minimum acceptable offer when stripped of norms. When making this guess, subjects had a pure incentive to make an accurate guess.

There are two main implications of this analysis: First, caution aimed at avoiding bargaining failure cannot explain our finding. If this was the case, larger groups would be good at avoiding bargaining failure, but not at actually guessing what the other side would

accept. As it turns out they are, and these guesses partially drive individuals' votes. Second, our finding is not just picking some tendency to blindly follow a behavioral norm. Instead, groups (especially democratic dyads) do better, in part, because they hold more accurate beliefs about the threshold for bargaining failure. These proposers' beliefs may, of course, include knowledge of norms that drive responders' demands, but they appear to also include other knowledge as well.

Appendix Figure A3: Accuracy of Proposers' Beliefs about Responders Minimum Acceptable Offer



Extended Literature Review of When Groups are *Not* Wiser

On balance, the scholarly literature on group versus individual decision-making appears to support the widely held belief that groups reach more rational or "better" decisions than individuals because they balance out biases and catch errors (Davis 1992). For example, Bornstein and Yaniv (1998) study groups versus individuals in one-shot ultimatum

barraging and find that, compared to individuals, groups behave more rationally by demanding more for themselves in the role of the proposer and accept less in the role of responder. Similarly, Cox (2000) studies groups versus individuals in an investment game and find the groups return significantly smaller amounts, in line with payoff maximization. Binder and Morgan (forthcoming) study groups versus individuals in an urn problem and in a monetary policy experiment and find that groups perform on average 4% better than individuals.

Nevertheless, these results (and many others) not withstanding, the psychology literature is by no means clear on whether "more = better." There are many instances where groups do not perform better than individuals in behavioral games. For example, Cason and Mui (1997) study groups versus individuals in a dictator game and find that groups can be more other regarding than individuals, despite the expectations of rational utility maximization. Cason and Mui (1997) also find the opposite result as Borenstein and Yanav: that groups may not always outperform individuals in ultimatum bargaining. Cox and Hayne (2002) explore decision-making of groups versus individuals in common value auctions and find that groups are more affected by the "disadvantage" of information, suggesting that they can be less rational than individuals. Bone et al (1999) and Rockenbach et al (2001) investigate group decision-making under risk and find only weak evidence that groups perform better than individuals. Koch and Sutter (2005) study groups versus individuals in beauty contest game and find mixed results: groups do not perform significantly better than individuals after one round of play, but – in support of our mechanism – when the game is repeated, groups learn the dynamics of the game significantly faster and outperform individuals over time. Puncochar & Fox (2004) investigate students' performance and accuracy in more real world course-related material and find that groups are indeed more accurate on exams than individuals, but at a cost of increased confidence in wrong answers.

These mixed findings suggest that there are other factors that may attenuate the advantages of group decision-making. These "scope conditions" are important for any theory of collective intelligence from group decision-making, and they complement the theoretical and empirical literature we have already cited on the limitation of a general Wisdom of the Crowds (Page 2008, Hong and Page 2004, 2009, 2012). Over time, however, many of these scope conditions have been more fully identified in experimental settings. Kerr, MacCoun & Kramer (1996) examine the relative susceptibility of individuals and groups to systematic judgmental biases and find that the relative magnitude of individual and group bias depends upon several factors, including group size, initial individual judgment, the magnitude of bias among individuals, the type of bias, and most of all, the group-judgment process.

Perhaps the biggest challenge to the conventional wisdom of group superiority is that groups can "attenuate, amplify or simply reproduce the judgmental bias of individuals (Kerr et al. 1996; 693). First, evidence suggests that group conformity and self-censorship may lead to "group-think", particularly in groups with designated leaders (Sniesek 1992, Kleindorfer et al. 1993; Mullen et al. 1994). Second, groups can polarize attitudinal judgments in many different settings (Davis 1992; Kerr et al, 1996; Cason and Mui 1997), although there is considerable evidence that groups also shift to moderate positions (Mosocvici 1985).

Importantly, however, many of the conditions that harm or attenuate the wisdom of crowds do not apply in our experimental setup; and, there is also reason to believe that

democratic decision-making is less vulnerable to many of these harmful conditions. A fuller exposition of this second argument can be found in the manuscript, where we describe the empirical differences between democracies and autocracies. Here, we'll simply note that in many of the cases above, the wisdom of the crowds argument breaks down because a single individual dominates decision making, as in Sniesek (1992), Kleindorfer et al. (1993), and Mullen et al. (1994). In our experiment, we avoided this by making people vote independently. Again this is not to assert that all decision-makers are completely independent in any real-world decision. Instead, to the extent individual inputs are independent, our treatment manipulation captures it. As we point out above, there is quantitative evidence that democracies institutionally have more points of independent influence, and that these "veto players" still influence crisis bargaining to some degree. Qualitatively, it also appears that crisis bargaining in democracies includes more independent voices. Meanwhile autocracies may include many human decision-makers, but compared to democracies, the qualitative evidence suggests that these actors are more likely to function as "yes men". Quantitatively, there are fewer actors who are institutionally independent.

Similarly, polarized judgment, which has harmed group decision-making in some experiments, could not readily occur in our experiment, as there was no opportunity for polarizing dynamics prior to players bargaining decision. While it might occur in the real world, there is no reason to believe it is more pernicious in democracies. Ideological factions can readily emerge in autocracies, and in the case where they do, members of the dominant faction may be less compelled to consider or incorporate alternative points of view.

Distinguishing distinct types of decision tasks also reveals at least one mechanism through which groups may outperform individuals. As Kocher and Sutter (2005) note "typically, groups perform better than individuals on (non-interactive) intellective tasks…meaning groups more often guess correctly than individuals (Hastie 1986; Levine and Moreland 1998), whereas "on judgmental tasks, there does not seem to be a significant difference" when there is no clear ex-post evaluation criterion for the quality of the judgment. Both real world crisis bargaining and ultimatum bargaining in our laboratory setting much more closely resemble intellective tasks, where is a clear ex-post evaluation of the quality of the judgment in so far as offers meat a minimum reservation price of an opponent or else they trigger bargaining failure.

Supplementary Information References

Austen-Smith, David, and Jeffrey S Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90(1): 34–45.

Canes-Wrone, Brandice, William G Howell, and David E Lewis. 2008. "Toward a Broader Understanding of Presidential Power: A Reevaluation of the Two Presidencies Thesis." *The Journal of Politics* 70(01): 1–16.

Chuah, S H et al. 2007. "Do cultures clash? Evidence from cross-national ultimatum game experiments." *Journal of Economic Behavior & Organization* 64(1): 35–48.

Coricelli, Giorgio, and Rosemarie Nagel. 2009. "Neural correlates of depth of strategic reasoning in medial prefrontal cortex." *Proceedings of the National Academy of Sciences* 106(23): 9163–9168.

Epley, Nicholas et al. 2004. "Perspective Taking as Egocentric Anchoring and Adjustment.." *Journal of Personality and Social Psychology* 87(3): 327–339.

Epley, Nicholas, Eugene Caruso, and Max H Bazerman. 2006. "When perspective taking increases

taking: Reactive egoism in social interaction.." *Journal of Personality and Social Psychology* 91(5): 872–889.
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): 361–383.
Henisz, Witold J. 2000. "The institutional environment for economic growth." *Economics & Politics* 12(1): 1–31.
Howell, William G, and Jon C Pevehouse. 2007. "When Congress Stops Wars: Partisan Politics and Presidential Power." *Foreign Affairs* 86(5): 95–107.
Howell, William G, and Jon C Pevehouse. 2011. *While Dangers Gather*. Princeton University Press.
Howell, William, and Jon Pevehouse. 2005. "Presidents, Congress, and the Use of Force." *International Organization* 59(01): 209–232.
Krueger, Joachim I., and Melissa Acevedo. 2008. "A Game-Theoretic View of Voting." *Journal of Social Issues* 64(3): 467–485.
LeVeck, Brad L et al. 2014. "The role of self-interest in elite bargaining." *Proceedings of the National Academy of Sciences of the United States of America* 111(52): 18536–18541.
Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. 1st ed. Princeton University Press.
Rand, David G. et al. 2013. "Evolution of fairness in the one-shot anonymous Ultimatum Game." *Proceedings of the National Academy of Sciences of the United States of America* 110(7): 2581–2586.
Tsebelis, George, and Seung-Whan Choi. 2009. "Democratic Peace Revisited: It is Veto Players." *Presented at the American Political Science Association Conference in Toronto*.
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): 559–573.