Democratic Reputations in Crises and War
Supplementary Appendix

Contents

A   Recruitment protocol

A.1   Recruitment letter to Knesset Members

הנך מוזמן להשתתף במחקר בחסות פרופסורים מאוניבריטאות פרינסטון, הרווארד, וסטנפורד שבארצות הברית.

פרויקט זה בוחן כיצד מקבלי החלטות חושבים על מדיניות חוץ. השאלון מיועד לנשים וגברים כאחד.

נהלי המחקר: המחקר כרוך בהשלמת שאלון אמריקאי ויקח כ "! דקות להשלים. שאלון זה מופץ לכל חברי הכנסת בישראל בהווה או בשני העשורים האחרונים. בכוונתנו לערוך בעתיד מחקר השוואתי עם חברי הפרלמנט הבריטי ואליטות בארצות הברית.

סודיות: מחקר זה הינו אקדמי בלבד, ועל מנת לשמור על אמינותו המחקרית אנו מחויבים לשמר את האנונימיות של המשתתפים. התשובות שלך יהיו אנונימיות לחלוטין ללא אפשרות לקשר את זהותך לתשובות שאתה מספק.

> להתחיל את השאלון, אנא הקש על הקישור הבא: ████████████
>
> אתה תתבקש להזין סיסמה בעלת ! ספרות. סיסמה זו הינה אישית,סודית, ונועדה לשימושך הפרטי בלבד. הסיסמה שלך היא: XXXXX
>
> בשל הפורמט של השאלון, אנו מבקשים שלא תענה על השאלון מהמכשיר הסלולרי שלך.(אם אתה מעדיף שנשלח לך עותק קשיח של השאלון באי מייל או בדואר, נשמח לעשות זאת!)

אנו מודים לך על על השתתפותך במחקר זה. אנו יודעים שזמנך יקר מאוד, ומעריכים את תרומתך למחקר בנושאים חשובים אלו. נשמח לשתף אותך בתוצאות המחקר לאחר שנעבד את הנתונים.

אם יש לך שאלות אנא פנה לפרופ' קרן ירחי מילוא מאוניברסיטת פרינסטון באימייל:
████████████ .

בברכה,

קרן ירחי מילוא,   יהושוע דויד קרצר,   מייקל טומז,
אוניברסיטת פרינסטון   !

Figure 1: Recruitment Letter

A.2   Recruitment procedures

We began the recruitment process for the elite sample by compiling a dataset of all 408 individuals who had served as members of the Parliament of Israel (i.e., the Knesset) from the beginning of the 14th Knesset in June 1996 through the 20th Knesset (the current Knesset) that was sworn in in

March 2015. We compiled a data set that included the following information about our population:

1. full name

2. party affiliation while in Knesset

3. names of all Knesset committees on which (s)he served

4. number of terms served

5. whether (s)he served as a minister in the government, and if so, what portfolios (s)he held

6. whether (s)he was a member of the Cabinet

Contact information for our participants was obtained through a variety of channels, including the Secretary of the Knesset, the Knesset Channel, the different parties' leadership offices in the Knesset and other government offices where former Knesset members are currently employed. Email addresses for all current members of the Knesset were obtained through the Secretary of the Knesset. To verify whether the contact information we obtained was correct, we either called or emailed all the former Knesset members from the last twenty years and asked them if they would be interested in taking a "10 minute electronic survey by a team of professors from leading American Universities." 29.4% of the initial population was removed from the sampling frame at this stage, either because the members were deceased, were too sick to participate, or because their contact information was out of date and newer contact information could not be found. This process left us with a sample of 288 potential candidates to take our survey. This pool included all 120 current members of the Knesset along with 168 former members whose contact information was available.

On July 10, 2015, we executed a soft launch of our on-line survey. The survey included a recruitment email, written in Hebrew (reproduced in Appendix §A.1), a link to our on-line survey, and an individual six-digit password that was pre-assigned to each member. In the following days, we emailed the invitation to all current and former members in our dataset. A few weeks later, we sent a reminder email to those who had not responded to the survey. We sent a third round of reminders a few weeks later. In between these rounds, we phoned former and current Knesset members or their assistants to remind them to take the survey. In early August, the Director of Academic Affairs at the Knesset, together with the Secretary of the Knesset, sent an email to all current Knesset members encouraging them to take the survey, repeating essentially the same information we provided in the introductory email.

In addition to the on-line survey, we created identical hard-copy versions of our survey. In mid-August we sent those who had not responded to our survey a reminder email and attached an electronic copy of our survey that could be opened in Microsoft Word. Respondents were given the option of either faxing or emailing the completed survey back to us. That same six-digit code was the only identifying information on the paper copies of the survey, allowing us to track completion among our sample population. Members of our research team also traveled to the Knesset on four separate occasions to invite current members to participate.

The entire recruitment process was done in Hebrew. Two Hebrew-speaking research assistants and one member of the research team who is a native Hebrew speaker corresponded with the members of the Knesset or their assistants. Participants were informed that there would receive no financial reward for taking the survey, but that we would be happy to share with them the results of the survey. Moreover, participants were promised full anonymity: with the exception of the research team, participants were assured that identifiable information would not be released or reported.

A.3 Participant verification protocol

We took several steps to increase our confidence that the current and former decision-makers participated in the study rather than members of their staff. First, in the introductory email we explicitly indicated that the questionnaire should be fielded by the decision-maker himself, and not by members of his or her staff. We explained that the code we provided to access the on-line survey was personal, and should not be shared with others. Importantly, we did not offer any material incentives for filling out the survey, to dissuade decision-makers and assistants for taking the survey for those material reasons.

Second, in the survey itself we asked the participants to enter their complete date of birth. This allowed us to compare this information with the date of the decision-makers in official Knesset records. Third, for the 75% of our sample consisting of former Knesset members, a Hebrew-speaking research assistant and one of the authors were both in touch with the decision-maker directly via phone or email, and confirmed with him/her that they were the ones taking the survey. Anecdotally, our research team found that many of our participants were quite eager for the opportunity to opine on issues of foreign policy to an outside audience.

In the case of some current Knesset members, after receiving approval from their parliamentary assistant, a Hebrew-speaking research assistant from our team or one of the authors gave the Knesset

members the survey directly and picked it up from them within a two-hour window. However, some Knesset members wished to maintain their anonymity and thus were not in direct contacted with the research team.

Finally, although we follow best practices, as is always the case with elite experiments, we should note that decision-makers who wished to "cheat" and delegate their participation to others could have probably found ways to do so. However, the combination of the types of questions asked in the survey, the absence of material compensation for survey completion, our explicit request the survey not be filled out by others, and the enthusiastic response to our survey from most of the decision-makers who took the survey leave us confident that the vast majority of them participated directly.

B   Study instrumentation

B.1   Experimental protocol (translated to English)

There is much concern these days about the spread of conflict. We are going to describe a situation the international community could face in the future. For scientific validity, the situation is general, and is not about any specific countries in the news today. Some parts of the description may strike you as important; other parts may seem unimportant. After describing the situation, we will ask you to make predictions about what you think will happen.

Here is the situation:

- Two countries are currently involved in a public dispute over a contested territory. The dispute has received considerable attention in both countries, because of the risk that disputes like these can escalate to the use of force.

- Country A [is a democracy/is a dictatorship]. Country B [is a democracy/is a dictatorship].

- Both countries have moderately powerful militaries, with large armies, moderate sized-navies, and well-trained air forces.

- Neither country is a close ally of the United States.

- Country A is slightly larger than Country B, though their economies are approximately the same size.

- Country A has moderate levels of trade with the international community. Country B has high levels of trade with the international community.

- The last time the two countries were involved in an international dispute, different leaders were in power.

Given the information available, what is your best estimate about whether Country A will stand firm in this dispute, ranging from 0% to 100%? [scaled from 0 (A will definitely **not** stand firm) to 100 (A will **definitely** stand firm)]

Now, we'd like to ask you a different question about the same situation. As a reminder, here is the situation:

- Two countries are currently involved in a public dispute over a contested territory. The dispute has received considerable attention in both countries, because of the risk that disputes like these can escalate to the use of force.

- Country A [is a democracy/is a dictatorship/recently transition to democracy]. Country B [is a democracy/is a dictatorship/recently transition to democracy

- Both countries have moderately powerful militaries, with large armies, moderate sized-navies, and well-trained air forces.

- Neither country is a close ally of the United States.

- Country A is slightly larger than Country B, though their economies are approximately the same size.

- Country A has moderate levels of trade with the international community. Country B has high levels of trade with the international community.

- The last time the two countries were involved in an international dispute, different leaders were in power.

If the dispute were to escalate and war were to break out, what is your best estimate about whether Country A will win, ranging from 0% to 100%? [scaled from 0 (A will definitely **not** win) to 100 (A will **definitely** win)]

B.2 Individual difference measures

We collected a range of demographic and dispositional characteristics measures across our seven samples. For our Israeli samples (both the elite sample of MKs, and the two public samples) participants completed questionnaires measuring military assertiveness, political ideology, stance on to the Arab-Israeli conflict, and international trust.[1] Military assertiveness (Herrmann, Tetlock and Visser, 1999) classifies individuals along a single "hawk-dove" dimension and has long been considered a key dimension of foreign policy beliefs. Recent work has also established its continued relevance, as military assertiveness remains as related to the decision to use force, even among a sample of U.S. leaders, as it was during the Cold War (Herrmann and Keller, 2004; Herrmann, Tetlock and Diascro, 2001; Kertzer and Brutger, 2016). This construct is equally predictive of foreign policy preferences in non-U.S. samples (see, e.g., Hurwitz, Peffley and Seligson 1993, Bjereld and Ekengren 1999 and Reifler, Scotto and Clarke 2011). Our political ideology item asks subjects to classify themselves along a single dimensions from "left" to "right" in politics, while a separate measure asks subjects to do the same with respect to the Israeli-Arab conflict. Our international trust measure is adapted from work by Brewer et al. (2004), which finds that generalized trust of other countries in the international system helps to structure beliefs about the foreign policy arena. The US sample includes these same items, apart from the Arab-Israeli conflict measure; the Korean and UK samples include the political ideology measure, but not the military assertiveness, international trust, or Arab-Israeli conflict measures.

B.2.1 Military assertiveness

1. The best way to ensure peace is through military strength

2. The use of force generally makes problems worse

3. Rather than simply reacting to our enemies, it's better for us to strike first.

All scaled from 1 (strongly agree) to 5 (strongly disagree). Items 1 and 3 reverse-coded such that higher scores indicated higher level of military assertiveness.

---

[1]Beliefs related to international trust were assessed only in *Public I* and the Knesset sample.

B.2.2 International trust

1. Some people say that Israel can trust other nations, while others think that Israel can't be too careful in dealing with other nations. Where would you place yourself on this scale from 1 (Israel can count on other countries) to 7 (Israel cannot count on other countries)?

Reverse-scored such that higher scores indicate more higher levels of trust.

B.2.3 Right-wing ideology

1. There is much talk of "left" and "right" in politics. How would you rate yourself on a left-right scale, from 1 (right) to 7 (left)?

Reverse-scored such that higher scores indicate more right-wing ideology.

B.2.4 Arab-Israeli conflict

1. How would you rate yourself on a scale from "left" to "right" with respect to the Israeli-Arab conflict where 1 one is on the far right and 7 is on the far left?

Reverse-scored such that higher scores indicate more right-wing ideology.

B.3   Secondary outcome measures

In addition to the primary outcome measures described in Appendix B.1, we also included a range of secondary outcome measures in one of our Israeli public samples, immediately following the administering of the primary outcome measures. The secondary measures were presented in a mixed factorial design, which included both between- and within-subject manipulations. First, the measures were presented in three blocks: one consisting of general beliefs about Country A, one soliciting beliefs about the consequences if Country A were to win the crisis/war, and one soliciting beliefs about the consequences if Country A were to lose the crisis/war. The order of the win/lose blocks was randomly assigned, to avoid potential order effects. Moreover, participants were either given the win/lose blocks in reference to the crisis scenario, or the war scenario. Asking questions about both types of scenarios as a between-subject factor allows us to reduce the length of the study instrumentation, while including questions about both positive and negative outcomes allows us to construct within-subject measures of the effect of losing on leader-level costs of war, for each participant.

1. What proportion of the citizens do you think believes force should only be used as a last resort?

    (1) 0-20% (2) 20-40% (3) 40-60% (4) 60-80% (5) 80-100%

2. What proportion of the national decision-makers do you think believes force should only be used as a last resort?

    (1) 0-20% (2) 20-40% (3) 40-60% (4) 60-80% (5) 80-100%

3. How sensitive do you think its citizens would be to casualties?

    (1) Very sensitive (2) sensitive (3) somewhat sensitive (4) not very sensitive

4. How sensitive do you think its citizens would be to the financial costs (e.g., increased taxes) of fighting?

    (1) Very sensitive (2) sensitive (3) somewhat sensitive (4) not very sensitive

5. If Country A initiated the dispute, how likely do you think it is that it will prevail in the dispute?

    (1) Very unlikely (2) unlikely (3) neither unlikely nor likely (4) likely (5) very likely

6. If the leader of Country A makes a public threat, how likely do you think it is that they'Äôll follow through?

    (1) Very unlikely (2) unlikely (3) neither unlikely nor likely (4) likely (5) very likely

7. How likely do you think it is that other countries would come to Country A's defense?

    (1) Very unlikely (2) unlikely (3) neither unlikely nor likely (4) likely (5) very likely

8. How well-trained do you think the soldiers of Country A are?

    (1) Very well trained (2) somewhat well trained (3)neither well trained nor poorly trained (4) somewhat poorly trained (5) very poorly trained

9. How strong do you think the morale of the soldiers is in Country A?

    (1) Very strong (2) somewhat strong (3) neither strong nor weak (4) somewhat weak (5) very weak

[Thinking back to the original crisis, if Country **A** were to back down. . .][If Country **A** were to lose the war. . .]

10. What effect do you think it would have on public support for the government?

    (1) Significantly increase (2) somewhat increase (3) not affect (4) somewhat decrease (5) significantly decrease

11. What effect do you think it would have on the likelihood of the government remaining in office?

    (1) Significantly increase (2) somewhat increase (3) not affect (4) somewhat decrease (5) significantly decrease

12. What do you think is likely to happen to the leader?

    (1) Remain in power (2) voted out of power (3) removed by a coup (4) exiled (5) death

[Thinking back to the original crisis, if Country **A** were to win. . .][If Country **A** were to win the war. . .]

13. What effect do you think it would have on public support for the government?

(1) Significantly increase (2) somewhat increase (3) not affect (4) somewhat decrease (5) significantly decrease

14. What effect do you think it would have on the likelihood of the government remaining in office?

(1) Significantly increase (2) somewhat increase (3) not affect (4) somewhat decrease (5) significantly decrease

For the analysis in Figures 3-4 of the main text, we convert these secondary outcome measures onto a probability scale to make the effect sizes directly comparable and substantively interpretable, but as a robustness check we replicate the analysis using the untransformed measures, and the same conclusions hold, as illustrated by the results in Appendix C.8.

C Supplementary analysis

C.1 Knesset results table

Table 1: Regression model: Knesset sample

	Resolve		Military Effectiveness	
	(1)	(2)	(3)	(4)
Democracy	−5.886	−7.280*	14.279***	13.318***
	(4.324)	(4.308)	(3.145)	(2.927)
Foreign affairs experience		4.917		6.062*
		(4.986)		(3.385)
Military assertiveness		−0.429		28.229**
		(18.695)		(12.698)
International trust		−16.469*		5.566
		(9.787)		(6.582)
Combat experience		−7.113		0.632
		(4.547)		(3.069)
L-R ideology		0.531		−27.748**
		(18.971)		(12.767)
Hawk (Arab/Israeli conflict)		−8.381		9.959
		(20.459)		(13.980)
Constant	59.136***	70.803***	44.302***	29.676***
	(3.057)	(10.514)	(2.224)	(7.117)
N	88	82	86	80
Adjusted R^2	0.010	0.021	0.187	0.252

**p < .05; ***p < .01

C.2 Israeli public samples

The first study of the Israeli public (*Israel I*) was piloted on September 30, 2015 and fielded from October 6-9, 2015. The second study of the Israeli public (*Israel II*) was piloted on January 17, 2016 and fielded from January 18-25, 2016. Descriptive statistics for each of these samples can be found in Table 2. Descriptive statistics for the Korean, and UK, and US samples are presented in the text.

Table 3 compares our three Israeli samples along several dimensions, all scaled from 0 to 1 for ease of interpretability. First, we note that current and former Knesset members do not differ very much on observed covariates; the only statistically significant difference between the two subsamples is that current leaders are on average 12 years younger than former leaders. The differences along the ideational dimensions are all small, and none of them are statistically significant. Unsurprisingly, our leader sample averages about 20 years older than our respondents from the Israeli public, and is generally less conservative, less hardline on the Arab-Israeli conflict, and more trusting of international institutions than the public at large.

Along the ideational dimensions that we measured, Israeli leaders are notably less conservative, less hawkish with respect to the Arab-Israeli conflict and more trusting of international institutions than the general public. For example, with respect to general ideology, Knesset members averaged a score of 0.45, a score that places them nearly an entire standard deviation *less conservative* than the mean score in the Israeli public samples. With respect to the Arab-Israeli conflict, our leaders were an entire standard deviation below the mean for the public samples. They also scored higher on international trust, and lower on military assertiveness, though those differences were not significant.

Table 2: Israeli Public Samples

	Israel I	Israel II
Male	53%	52%
Education:		
No High School degree	2%	2%
High School degree	33%	26%
Some college	22%	22%
College degree	27%	31%
Masters degree	14%	17%
Doctoral degree	2%	2%
Military Experience:		
Did not serve	22%	20%
Served, no active combat	50%	50%
Combat experience	28%	31%
Location:		
Jerusalem	10%	10%
Tel Aviv	12%	14%
Central Zone	18%	18%
Haifa	15%	14%
Northern Region	11%	13%
Southern Region	12%	9%
Lowland	10%	11%
Sharon Area	8%	9%
Yehuda and Shomron	3%	2%
Religiosity:		
Secular	60%	54%
Traditional	19%	30%
Religious	15%	14%
Orthodox	5%	3%
Birth Country:		
Israel	81%	81%
Former USSR	10%	10%
Other	9%	9%
	Mean (SD)	Mean (SD)
Age	41.8 (14.6)	41.8 (14.7)
Military Assertiveness	0.58 (0.19)	0.56 (0.20)
Right Wing Ideology	0.61 (0.25)	0.59 (0.23)
Hawkishness (Arab-Israeli conflict)	0.63 (0.25)	0.62 (0.25)
International Trust	0.32 (0.27)	—

	Knesset			Israeli Public		Elite–Public Gap	
	Current	Former	Overall	*I*	*II*	*I*	*II*
Military Assertiveness	0.61	0.61	0.61	0.58	0.56	0.03	0.05
Right Wing Ideology	0.47	0.44	0.45	0.61	0.59	**−0.16**	**−0.14**
Hawkishness (Arab/Israeli)	0.44	0.37	0.39	0.63	0.62	**−0.24**	**−0.23**
International Trust	0.37	0.41	0.40	0.32	—	**0.08**	—

Table 3: Elite-Mass Comparison: Statistically significant differences in means between public sample and leader (overall) sample depicted in **bold**; p-values calculated via Wilcoxon rank-sum tests.

C.3   Israeli public results table

Table 4: Regression models: Israeli public samples

	Resolve				Military effectiveness			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Democracy	−4.591***	−4.752***	−1.675	−1.857	4.964***	4.984***	8.227***	7.948***
	(1.232)	(1.237)	(1.273)	(1.281)	(0.985)	(0.986)	(1.114)	(1.118)
Male		−2.174		−3.176**		−1.366		−1.947
		(1.405)		(1.449)		(1.119)		(1.265)
Age		0.035		0.028		0.071*		0.086**
		(0.047)		(0.047)		(0.037)		(0.041)
Education		−3.825		0.258		−2.128		−6.630***
		(2.760)		(2.877)		(2.198)		(2.512)
Political knowledge		5.028		6.370*		−2.491		1.040
		(3.635)		(3.540)		(2.896)		(3.088)
Combat experience		0.717		−1.339		−0.980		−0.679
		(1.509)		(1.530)		(1.202)		(1.335)
Military assertiveness		3.803		0.781		9.819***		7.002**
		(4.051)		(4.048)		(3.227)		(3.532)
International trust		−0.780				0.034		
		(2.474)				(1.970)		
Ideology		5.323		2.523		1.694		6.681
		(4.452)		(4.781)		(3.546)		(4.171)
Hawk (Arab–Israeli)		−1.474		2.239		−1.630		−3.282
		(4.334)		(4.552)		(3.452)		(3.977)
Born in Israel		0.118		3.355**		−0.239		−0.029
		(1.631)		(1.682)		(1.299)		(1.467)
Constant	57.864***	50.825***	59.913***	49.681***	48.523***	43.914***	50.062***	44.474***
	(0.876)	(4.701)	(0.897)	(4.379)	(0.701)	(3.745)	(0.784)	(3.825)
N	1,100	1,089	1,111	1,102	1,100	1,089	1,110	1,101
Adjusted R²	0.012	0.013	0.001	0.007	0.022	0.032	0.046	0.060
Sample #	I	I	II	II	I	I	II	II

**p < .05; ***p < .01

C.4 Heterogeneous effects in the Israeli sample

Our main approach in the paper to test whether the results generalize outside of the Israeli context is to field the study in a range of other democratic countries (the United States, South Korea, and the United Kingdom), where we find broadly similar results.

We also designed the Israeli vignettes with questions of generalizability in mind. For example, the repeated conflicts in which Israel has been involved might cause Israelis to see their adversaries as more resolved than participants in other countries, a particular danger if those subjects also interpreted the scenario as being about Israel. To preempt this, we designed our studies to invoke two hypothetical countries with features that are inconsistent with those of Israel: only half the participants see a country described as a democracy, and that democracy is described as *not* being a close ally of the United States. Yet even if participants did have Israel in mind, it is unlikely to change the interpretation of our findings. For example, if Israelis simply see every adversary as more resolved (than a Canadian respondent would, for example), this might explain why we see a negative effect of democracy on beliefs about resolve in crises, but would not explain the positive effect of democracy on beliefs about resolve in military battles.

Similarly, Israeli participants might be more likely to interpret the scenario in the context of the Middle East, while participants in other regions might be more likely to rely on other analogies or mental models instead (Khong, 1992). To test whether our findings are contingent on the mental models participants employ, we presented respondents in one of the Israeli samples with a list of potential regions (Central Asia, North America, Western Europe, Sub-Saharan Africa, Middle East/North Africa), and asked them to rank the regions in terms of their likelihood of experiencing the scenario described in the experiment. In this manner, we can assess whether participants who imagine the scenario taking place in different parts of the world understand the relationship between democracies, resolve, and war differently. We employ Yilmaz, Aslam and Robertson's (2008) τ_{ap} statistic, a weighted rank similarity measure,[2] to test whether participants who imagine the scenario taking place in different regions make different predictions about the relationship between democracy, crises, and war. Reassuringly, we find no evidence of heterogeneous treatment effects across rank similarity scores, indicating that the regional context participants have in mind does not affect the

---

[2]We assigned each participant a rank similarity score in reference to a ranking consisting of the average position of each region. Unlike unweighted rank similarity statistics like Kendall's τ_b, weighted rank similarity measures, commonly used in information science, assign differential penalties based on the position in the ranking at which they occur. We also replicate our analyses using Kendall's τ_b, and find the results hold.

conclusions we reach.[3] This suggests our findings are robust to the identities of the belligerents our respondents have in mind.

C.5 Downsampling Israeli public results

Figure 2: Downsampled crisis results
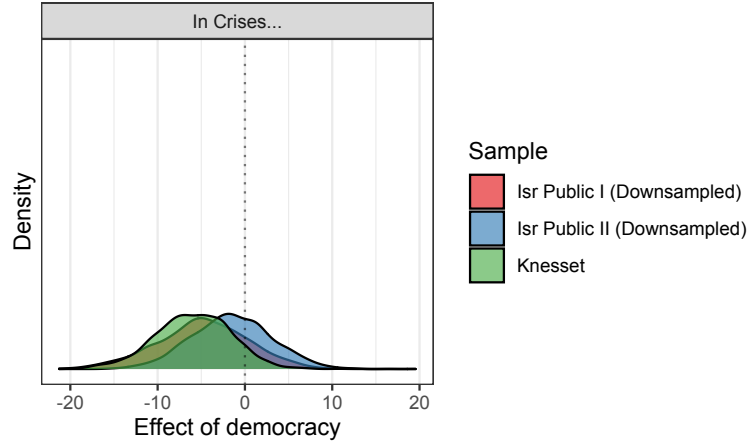
The figure shows that the heavier tails in the distribution of treatment effects in the Knesset results in the main paper is due to the relatively small sample size rather than peculiarities about the sample; when we downsample the results from the two public samples, sampling $N = 88$ observations with replacement rather than $N = 1111$ or 1599, we obtain similar distributions as in the Knesset survey.

---

[3]More formally, we estimate a set of regression models, regressing each dependent variable on the democracy treatment, the regional similarity measure, the interaction between the two, and a variety of demographic controls. The interaction lacks statistical significance across all four models (on resolve: $p < 0.85$ for τ_{ap}, $p < 0.59$ for τ_b; on military effectiveness: $p < 0.70$ for τ_{ap}, $p < 0.49$ for τ_b).

C.6 Comparison of effect size magnitude with large-N work

The experimental results found fairly consistent patterns across all seven samples. One question that remains concerns the substantive size of the effects: how important is democracy in crises if participants only see it as having at most a 6% effect? How does this compare to the observed effect of democracy in crises as measured in large-N work?

To address this question, we use replication data that Downes and Sechser (2012) provide for Schultz's (2001) pathbreaking exploration of democratic performance in disputes. Table 5 presents four sets of models taken from models 3 and 4 from table 5.5 in Schultz (2001, 146-147). The data consist of bilateral militarized interstate disputes (MIDs) using version 2.1 of the MID data; because of missingness, the analysis focuses on the years 1816-1984.[4] The dependent variable is a dichotomous variable indicating whether the MID was reciprocated. The first two columns of Table 5 include data from the two world wars, while the second two do not; each model is estimated using either a logistic regression, or a linear probability model.

The centerpiece of the models is the first three rows, which indicate the regime type of the initiator and target in each dispute. Schultz is interested in particular in the coefficient estimate for the democratic initiator variable, which is negative and significant across all four models, suggesting that democracies fare better in crises because their targets "back down without a fight." (p. 9). The coefficients in columns 2 and 4 are already on a probability scale, but to interpret the substantive effects for the logit model, we calculate the predicted probabilities, presented in Table 6, which depicts predicted probabilities using models 1 and 3, for contiguous dyads, a non-democratic target state, and a policy demand, for with all other covariates set to their mean values for each of the four power configurations. The results suggest effect sizes ranging from -6.3% to -8.5%. In this sense, they are similar to the results from the linear probability models in columns 2 and 4, which depict effect sizes for democratic initiators facing non-democratic targets as ranging from -6.6% to -7.9%.

A direct comparison between Schultz (2001) and our experimental results is difficult, for any number of reasons: questions about the limitations of the MID data for testing these types of questions (Downes and Sechser, 2012), questions about the limitations of observational data more generally in causally identifying the effects of democracy (Tomz and Weeks, 2013), differences between the setup of our experimental scenario and the covariate profile in the MID data, and so on. Despite these caveats, and the different conclusions we draw from our data, it is notable that the

---

[4]For more details about the sources of the variables in the model, see Schultz (2001); Downes and Sechser (2012).

Table 5: The probability of crisis reciprocation: Bilateral MIDs

	(1) World wars included	(2) World wars included	(3) World wars excluded	(4) World wars excluded
Democratic Initiator	−0.314*	−0.066*	−0.374*	−0.079*
	(0.181)	(0.038)	(0.191)	(0.041)
Democratic Target	−0.003	−0.002	−0.051	−0.012
	(0.159)	(0.034)	(0.169)	(0.036)
Both Democratic	0.364	0.077	0.489	0.102
	(0.343)	(0.073)	(0.361)	(0.077)
Major Power Initiator-Major Power Target	−0.171	−0.036	−0.314	−0.067
	(0.285)	(0.061)	(0.306)	(0.065)
Major Power Initiator-Minor Power Target	−0.254	−0.054	−0.242	−0.054
	(0.201)	(0.043)	(0.210)	(0.045)
Minor Power Initiator-Major Power Target	0.218	0.045	0.232	0.051
	(0.252)	(0.054)	(0.263)	(0.057)
Initiator's share of capabilities	0.070	0.017	0.060	0.012
	(0.240)	(0.052)	(0.247)	(0.053)
Contiguous	0.501***	0.110***	0.547***	0.120***
	(0.145)	(0.031)	(0.155)	(0.033)
Alliance portfolio similarity	0.143	0.033	0.167	0.039
	(0.223)	(0.048)	(0.234)	(0.050)
Status quo evaluation of initiator	−0.134	−0.026	−0.083	−0.016
	(0.192)	(0.041)	(0.205)	(0.043)
Status quo evaluation of target	−0.219	−0.048	−0.261	−0.057
	(0.215)	(0.046)	(0.229)	(0.049)
Territory	0.288*	0.069*	0.273	0.063*
	(0.165)	(0.036)	(0.174)	(0.038)
Government	0.327	0.074	0.236	0.053
	(0.375)	(0.080)	(0.383)	(0.082)
Policy	−1.142***	−0.259***	−1.215***	−0.279***
	(0.149)	(0.032)	(0.158)	(0.034)
Other	−0.649	−0.151	−0.608	−0.141
	(0.569)	(0.127)	(0.588)	(0.132)
Constant	0.051	0.513***	0.075	0.521***
	(0.269)	(0.058)	(0.281)	(0.061)
N	1,305	1,305	1,153	1,153
AIC	1,646.909		1,447.088	
Model type	Logit	OLS	Logit	OLS
World war years dummies	Yes	Yes	No	No

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

Table 6: Predicted probabilities of crisis reciprocation

Initiator-Target	Nondemocratic initiator	Democratic initiator	Effect
World wars included			
Major Power - Major Power	30.8%	24.6%	-6.3%
Major Power - Minor Power	31.7%	25.4%	-6.4%
Minor Power - Major Power	39.7%	32.5%	-7.2%
Minor Power - Major Power	39.1%	31.9%	-7.2%
World wars excluded			
Major Power - Major Power	27.8%	20.9%	-6.9%
Major Power - Minor Power	32.2%	24.7%	-7.6%
Minor Power - Major Power	39.9%	31.4%	-8.6%
Minor Power - Major Power	39.3%	30.8%	-8.5%

Predicted probabilities calculated using the coefficient estimates from Table 5, columns 1 and 3. The predictions shown are for a contiguous dyad, a nondemocratic target state, and a policy demand, with all other variables set to their mean values within each power configuration.

effect sizes we find for democracies in crises in our experiment are similar to those found in one of the canonical large-N studies of this question. In this sense, our maximum effect size of democracy in crises of 6% is substantively meaningful in the context of other work that has also explored this question.

C.7 Exploiting the two-stage nature of the experimental design

As noted in the main text, one of the innovations of our experimental design is that we study respondents' beliefs about democracies in both crises and war, which not only broadens the scope of the investigation to democracies' reputations, but also more closely connects our research design to contemporary bargaining theories of war, which view crises and war as part of one continuous process (Smith and Stam, 2004; Powell, 2004). This also enables us to explore questions in the like the difference-in-difference between democracies vs dictatorships in crises vs wars: one of the striking patterns we show in Figure 4 in the main text is that across all seven samples, respondents believe democracies are at a significantly larger advantage in wars than in crises. There are two potential interpretations for this pattern of results. One holds that democracies are seen as being at an advantage in war *despite* disadvantages in crises — as would be the case if each phase implicates only a partially overlapping set of causal mechanisms; considerations of military effectiveness (where democracies have positive reputations) may be more salient in reputations in war, and considerations of the credibility of democratic threats may loom larger in reputations in crises, for example. The other holds that democracies are seen as being at an advantage in war *because* of their perceived disadvantages in crises. This would be consistent with rationalist theories in which democracies won't let crises escalate into war unless they're serious about winning.

To adjudicate between these competing interpretations, we carry out two different types of empirical tests. First, we exploit the two-dimensional structure of the experiments to test whether perceived democratic advantages in war outweigh perceived disadvantages in crises. Second, we estimate a series of average controlled direct effects to test for the causal relationship between the two sets of experimental results. We discuss each in turn.

First, we utilize a simulation-based approach to exploit the two-dimensional structure of our experiments. The top panel in Figure 3 presents a series of mosaic plots. Here, for each of our seven samples, we dichotomize each dependent variable at the 50% mark (dropping those respondents in each sample who provided answers right at 50%, such that crisis observations are coded either as standing firm or backing down, and war observations as either winning or losing). The x axis therefore presents the proportion of respondents who thought the state in question was more likely to stand firm versus back down in the crisis, while the y axis depicts the proportion of respondents who thought the state in question was more likely to win versus lose in the war; in each sample,

24

responses in the democracy condition are presented in red, and responses in the dictatorship condition in aquamarine. The mosaic plots show considerable variation across our samples in terms of the proportion of respondents who believe democracies are at an advantage in both crises and wars (with most respondents perceiving a democratic advantage in war across most samples, but more between-sample variation in beliefs about democratic performance in crises). To more clearly illustrate the interaction between different types of democratic reputations, then, we conduct a simple bootstrapping exercise, in which, for each sample, we:

- Create a simulated dataset by sampling from the actual data with replacement

- Using the simulated dataset, calculate the product of both dependent variables, expressed as a percentage (e.g. if a respondent saw the country as having a 50% chance of standing firm and a 60% chance of winning, the score would be 0.5 x 0.6=30%)

- Calculate the mean score for respondents in the democracy condition, and the mean score for respondents in the dictatorship condition

- Subtract the mean score for respondents in the dictatorship condition from the mean score for respondents in the democracy condition, to calculate the overall effect of democracy

- Repeat $B = 1500$ times to derive the bootstrapped distribution of scores

We plot these scores in the density plots in the bottom panel of Figure 3. They show that for one sample (USA), democracies' perceived disadvantage in crises outweighs their perceived advantage in war, another sample (Israel I), the two phases cancel each other out, and in the remaining five samples, democracies' advantage in war outweighs their perceived disadvantage in crises.

Second, to study the relationship between democratic reputations in crisis and war in a causal fashion, and provide some leverage on the question of whether reputations transfer across contexts, we turn to sequential G-estimation (Acharya, Blackwell and Sen, 2016). If democratic reputations in crises affect democratic reputations in war, estimating the effect of regime type on beliefs about a state's likelihood of winning war controlling for beliefs about the state winning in a crisis would induce post-treatment bias. Sequential G-estimation sidesteps this concern, enabling us to estimate the direct effect of democracy on perceived war outcomes, controlling for the resolve in crisis mediator without inducing post-treatment bias. Table 7 presents the results for all seven of our samples. The first column presents the average treatment effect (ATE) of regime type on perceived war outcome;

Figure 3: Exploiting the two-stage structure of the experiment
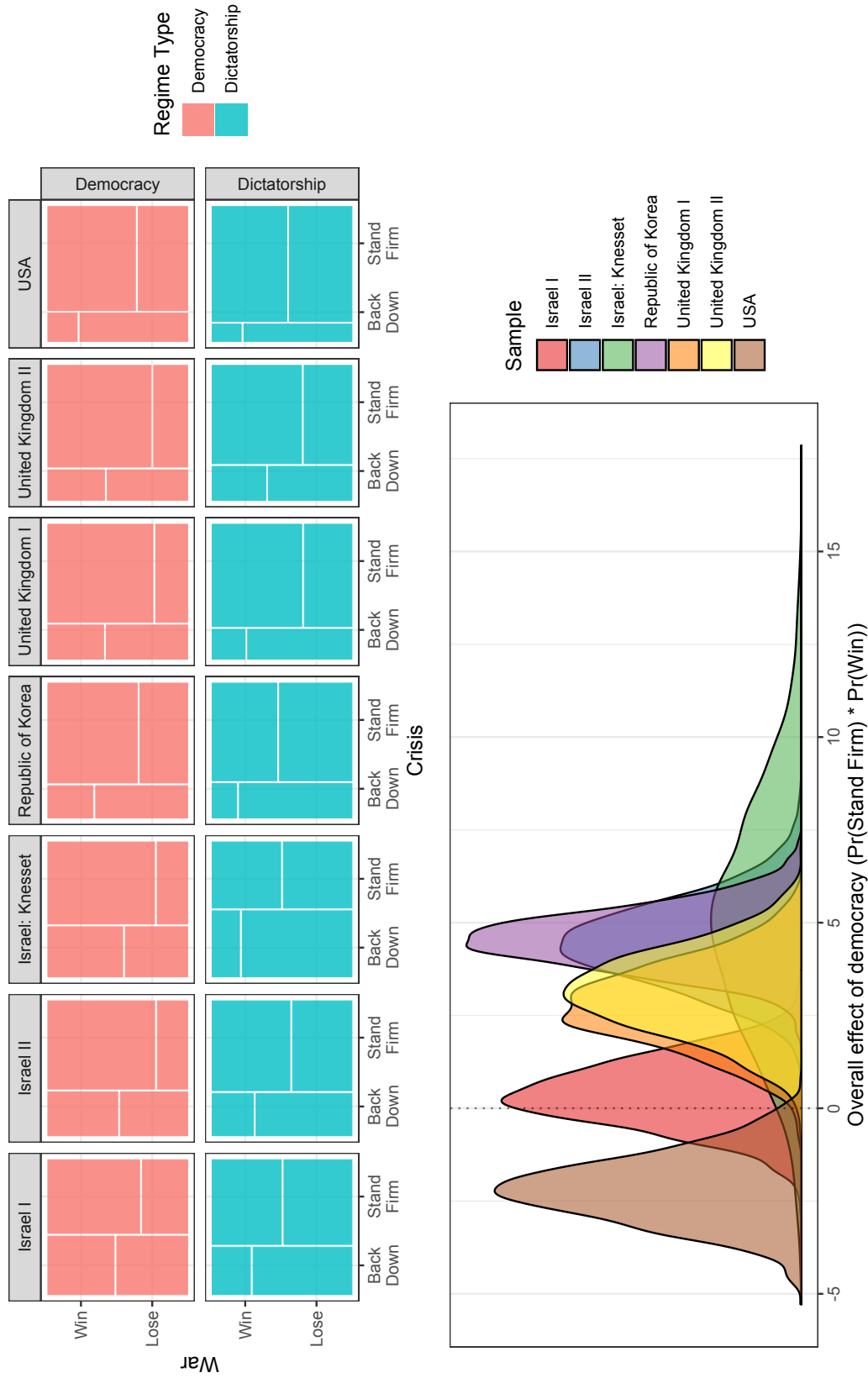
The top panel presents a series of mosaic plots, in which responses to each dependent variable within each sample have been dichotomized at 50%, to depict the proportion of each sample who perceives the state in question as being more likely to stand firm vs back down in a crisis (on the x axis), and win vs lose in war (on the y axis). The bottom panel presents the bootstrapped distribution of the overall effect of democracy, calculated as the product of the likelihood that the state in question will stand firm in a crisis, and win in a war. It shows that in general, democracies' advantages in war are perceived as outweighing any disadvantages in crises, although this varies across samples.

the second column presents the average controlled direct effect (ACDE) of regime type on perceived war outcomes controlling for perceived resolve in crises, estimated using sequential g-estimation. If perceived democratic disadvantages in crises caused greater perceived advantages in war, we should expect that controlling for the effect of democracy on the former should increase the effect of democracy on the latter. Instead, the results show that for four of the seven samples, the ACDE is substantively *larger* than the ATE: that is, controlling for democratic reputations in crises causes democratic reputations in war to become *more* positive. For the other three samples, the difference between the ATE and ACDE is negative, but relatively small in size. Altogether, then, these results are more consistent with democratic reputations in war being positive despite, rather than because, of perceived democratic reputations in crises, implying that democratic reputations in war would be even stronger if not for their perceived disadvantage in crises.

Table 7: Sequential g-estimate of effect of regime type on perceived war outcome

	ATE	ACDE
Israel I	4.964	5.833
	(0.985)	(0.954)
Israel II	8.227	8.743
	(1.114)	(1.045)
Israel Knesset	14.279	15.533
	(3.145)	(2.902)
Republic of Korea	6.318	6.313
	(0.783)	(0.754)
United Kingdom I	3.579	3.460
	(1.089)	(0.991)
United Kingdom II	3.671	3.432
	(1.055)	(0.967)
USA	4.061	4.342
	(1.499)	(1.419)

Note: the first column presents the average treatment effect (ATE) of regime type on perceived war outcomes, while the second column presents the average controlled direct effect (ACDE) of regime type on perceived war outcomes controlling for perceived resolve in crises. Standard errors shown in parentheses.

C.8 Effect of regime type treatment on secondary outcome questions

As noted above, for the analysis in Figure 3 of the main text, we convert the secondary outcome measures onto a probability scale to make the effect sizes directly comparable and substantively interpretable (the primary outcome measures are already on a probability scale, and thus don't need to be transformed). In Figure 4 we present the treatment effects for the untransformed secondary outcome measures. Importantly, the same conclusions hold.[5]

Figure 4: Untransformed outcome measures

(a) Democratic reputations in crises

(b) Democratic reputations in wars

---
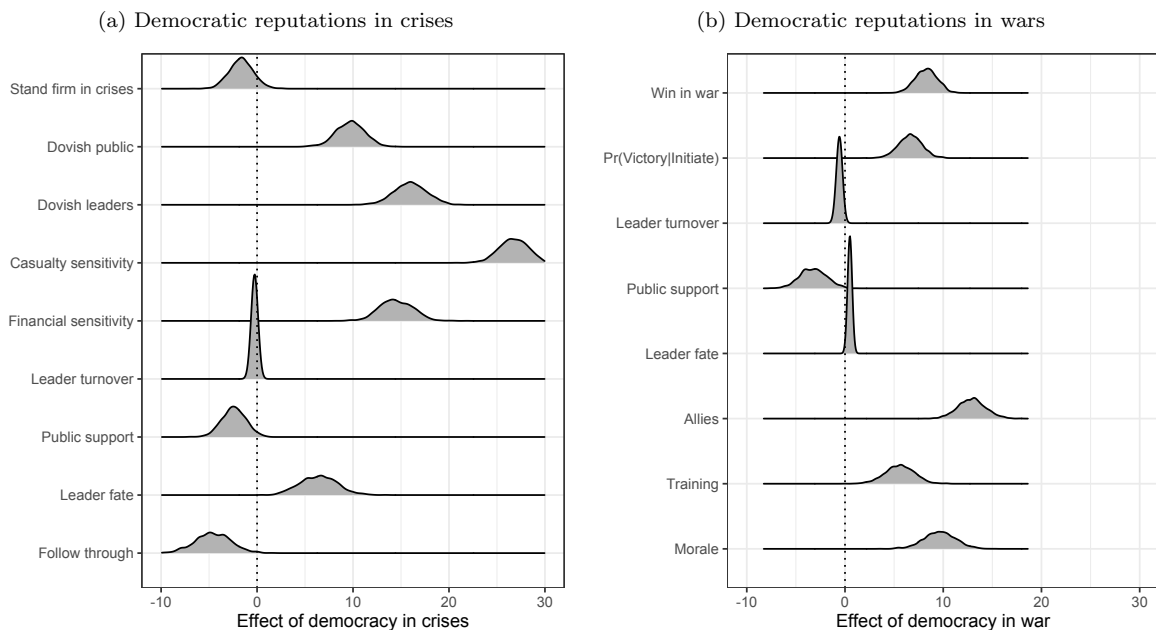
[5]The key difference is with the "Leader fate" variable, which because it is untransformed here, should be interpreted such that higher values indicate a more favorable anticipated fate for the leader; in Figure 3 in the main text, the variable is transformed to become an irregular exit variable (indicating the probability that respondents attribute to the leader leaving office through a coup, exile or death).

D   Experimental Design Considerations

D.1   Disentangling reputations from other concepts

In our paper, we build on a standard conception of reputation in IR offered by Dafoe, Renshon and Huth (2014, 372): "a belief about a trait or behavioral tendency of an actor, based on that actor's past behavior; usually widely held." As we note in the body of the paper, these reputations were most commonly studied as belonging to states and leaders and being about resolve (whether or not an actor would stand firm in future interactions). We attempt to broaden the study of reputations along both dimensions, taking our cue from other recent work in IR suggesting "reputations may adhere to any kind of agent: a leader, a ruling group, an organization. . . institutions, or any other factor that shapes a country's behavior. . . " (Renshon, Dafoe and Huth, 2018, 326). While this represents a departure from the narrower way in which reputation is traditionally modeled in IR, it is consistent with how reputations are studied in both psychology and other quadrants of political science, where American politics scholars use "party reputations" (e.g. Pope and Woon, 2009) to refer to the same thing that Comparative politics scholars call "party brands" (e.g. Lupu, 2016) to refer to the same thing that political psychologists refer to as "partisan stereotypes" (e.g. Kertzer, Brooks and Brooks, 2021). The question "do Republicans have a reputation for opposing tax increases?" is similar to "do democracies have reputations for backing down in crises?": both involve making inferences about an actor's traits or behavioral tendencies based on socially shared beliefs about the actor.

However, reputations might in some cases overlap with other concepts and it's worth considering how our design approaches these potential issues. One such concept is *identity*, which is used in a wide variety of ways in IR (e.g. Bially Mattern, 2001; Thies, 2002; Hayes, 2012). We follow Abdelal et al. (2009), who define identity as a "social category that varies along two dimensions–content and contestation." That content may be defined in several ways (Abdelal et al. 2009 offer a framework with four "types" of content, for example), but usually in IR we focus on *social* or *collective identity* (which define the individual in relation to groups) rather than *personal identity* (the extent to which individuals are defined in reference to other individuals) (Tajfel and Turner, 1986), and fundamentally social identity as a concept helps us understand norms that define groups, the goals of group members, and how those identities affect our own behavior and relational comparisons to outgroup members. Central to identity is the notion of *subjective identification*: we often study social identity

by understanding the categories actors use to understand who they are (Egan, 2020), as in experiments where researchers prime respondents to change their identity attachments (Levendusky, 2018; Powers, 2021).

Where identity and reputation are perhaps similar is that both involve some form of categorization, but although categorization is an important part of theories of identity (Tajfel and Turner, 1986), it is not coterminous with it (Rosch and Lloyd, 1978). Indeed, it is worth noting what our experiments do *not* do:

- Manipulate the social identity of our participants, as in traditional social identity theory (SIT) experiments (Powers, 2021)

- Manipulate the social identity categories of the actors in the experiment, as in Johns and Davies (2012), who vary whether a country is predominantly Muslim or Christian to test microfoundations of Huntington's clash of civilizations hypothesis

- Manipulate whether the countries in the dispute are ingroup or outgroup members (e.g. Rousseau and Garcia-Retamero, 2007) – as discussed in greater detail in section 2.4 of this memo, we deliberately manipulate regime type while controlling for social distance from the United States by designing the scenario so that neither country involved is an ally of the United States. And, in one of the Israeli experiments, supplementary analysis in Appendix C.5 shows that our results are robust to what region of the world respondents perceive the dispute as taking place (and thus presumably, the identities of the belligerents).

- Measure respondents' own degrees of subjective identification (Herrmann, Isernia and Segatti, 2009)

Instead, our study asks a more basic question: if we manipulate the level of democracy of a country in a dispute with another state, what inferences do respondents draw about the state's behavioral tendencies in crises in war? In this sense, our manuscript is perhaps closest to Bush and Zetterberg (2021), who study what they refer to as "reputations for democracy". Using experiments in Sweden and the United States, they manipulate whether a state has carried out certain behaviors, and ask respondents to draw inferences about how democratic they think the state is. In our case, we field experiments in the United States, Israel, South Korea, and the United Kingdom to study the inverse question: if you know a state is democratic, what inferences do you draw about its

behavioral tendencies in crises and war?

A related issue concerns the notion of "democracy as a signal." In order to function, signaling requires common knowledge: in the absence of socially shared understandings about the signal's meaning, the recipient will not draw the inferences the sender intends (Krauss and Morsella, 2000). In our case, there is no formal signaling in the experiment itself since the state isn't carrying out communicative actions itself,[6] but the common knowledge or socially shared understanding requirement remains: in order for respondents to draw inferences about the state's behavioral tendencies upon learning that the state is democratic, they must draw upon socially shared understandings of democracy's foreign policy implications.[7] Consistent with the psychological literature, we call these socially shared understandings reputations. If observers use a state's regime type as indicative of its future behavior in crises or wars, and their responses point in similar directions, this suggests that regime types have reputations: people carry around socially shared understandings about the foreign policy implications of regime type with them in their heads. The goal of our piece is to find out what they are.

D.2   Manipulating Regime Type and Alliance

In our experimental design, we manipulated regime type by randomly assigning respondents to descriptions of either a "democracy" or a "dictatorship." There were two primary considerations in this design choice. First, we focused on the main contrast between democracies and non-democracies for the same reason that earlier observational work did. Just as the literature on regime type in IR with observational data began by simply dichotomizing regime type into democracies versus non-democracies before making more nuanced distinctions, as the first piece to study democratic reputations in IR we begin by focusing on the more basic distinction between democracies and dictatorships, leaving more fine-grained distinctions for future work. Manipulating regime type in a rather broad manner also facilitates comparability with other experimental approaches in comparative and international politics, which nearly universally manipulate regime type in the same manner as we do (Mintz and Geva, 1993; Johns and Davies, 2012; Tomz and Weeks, 2013; Bell and Quek, 2018; Johns and Davies, 2019; Kertzer, Renshon and Yarhi-Milo, 2021). Given the novel nature of

---

[6]For examples of recent experimental work on signaling, see e.g. Quek (2016); Yarhi-Milo, Kertzer and Renshon (2018)

[7]Note that these understandings need to be socially shared in order for the treatment effect to be significant; if all respondents drew upon independent mental models, and the models were uncorrelated with one another, different respondents would reach different conclusions, cancelling one another out.

our study, and our interest in putting extant theories with implications for democratic reputations to the test, comparability with other work is a significant benefit of our approach.

The second consideration was statistical power. In one of the Israeli public samples and the US sample, we included a third category of regime type: a country that had recently transitioned from an autocratic regime to a democracy. As the cell means plotted in Figure 5 show, the results for this transitional democracy were in between the democracy/dictatorship categories (with results in crises closer to democracy than to dictatorship, and results in war more squarely in between the two), and thus statistical power considerations (particularly with the smaller Knesset sample) impelled us to simplify the design for our other studies.

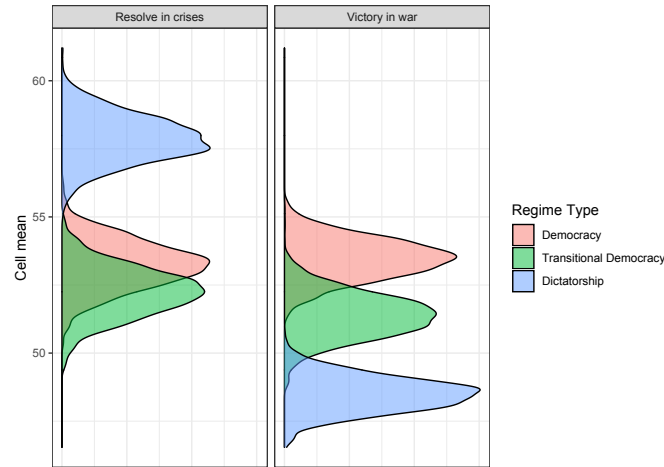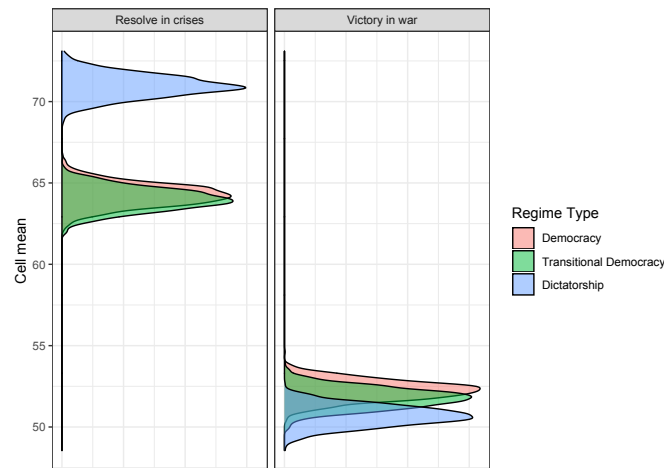Figure 5: Results with transitional democracy condition (Israeli public sample)

Figure 6: Results with transitional democracy condition (US sample)

A related issue concerns other details in the experimental vignette. In our design, we indicated that neither country in the dispute was an ally of the United States. We controlled for alliance ties in this manner in the experimental design in order to avoid two concerns in particular. The first is information leakage (Tomz and Weeks, 2013). A common critique of the literature on democracies in war is that our theories of democracy are just really theories of US hegemony in disguise, and that the salutary effects of democracy are simply the benefits that accrue to the US and its friends (Rosato, 2003). If we hadn't specified anything about each country's alliance ties, our concern would be that respondents would assume the country was a US ally if it was a democracy, and assume it wasn't an ally it was a dictatorship, leading to information leakage and bundled treatment problems; holding alliance ties fixed across both conditions avoids this concern. As noted in Section D.1, this allows us to sidestep potential identity-based confounds as well, as would be the case if respondents presumed the democracy was socially close to their own country, and the dictatorship was socially distant.

The second concern we sought to avoid was projection: although we deliberately told respondents that the scenario wasn't about their own country, we were still concerned some would think—as R2 expressed concern about—that we were simply projecting their own state into the scenario. This latter issue would be problematic insofar as it would likely trigger "positive illusions" (e.g., overconfidence) if subjects were to perceive the vignette as being about their own country (Brown, 1986).

References

Abdelal, Rawi, Yoshiko M. Herrera, Alastair Iain Johnston and Rose McDermott. 2009. *Measuring Identity: A Guide for Social Scientists.* Cambridge: Cambridge University Press.

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining causal findings without bias: Detecting and assessing direct effects." *The American Political Science Review* 110(3):512–529.

Bell, Mark S and Kai Quek. 2018. "Authoritarian public opinion and the democratic peace." *International Organization* 72(1):227.

Bially Mattern, Janice. 2001. "The Power Politics of Identity." *European Journal of International Relations* 7(3):349–397.

Bjereld, Ulf and Ann-Marie Ekengren. 1999. "Foreign policy dimensions: A comparison between the United States and Sweden." *International Studies Quarterly* 43(3):503–518.

Brewer, Paul R, Kimberly Gross, Sean Aday and Lars Willnat. 2004. "International trust and public opinion about world affairs." *American Journal of Political Science* 48(1):93–109.

Brown, Jonathon D. 1986. "Evaluations of self and others: Self-enhancement biases in social judgments." *Social cognition* 4(4):353–376.

Bush, Sarah Sunn and Pär Zetterberg. 2021. "Gender Quotas and International Reputation." *American Journal of Political Science* 65(2):326–341.

Dafoe, Allan, Jonathan Renshon and Paul Huth. 2014. "Reputation and Status as Motives for War." *Annual Review of Political Science* 17:371–393.

Downes, Alexander and Todd Sechser. 2012. "The Illusion of Democratic Credibility." *International Organization* 66(3):457–489.

Egan, Patrick J. 2020. "Identity as dependent variable: How Americans shift their identities to align with their politics." *American Journal of Political Science* 64(3):699–716.

George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences.* Cambridge, MA: MIT Press.

Hayes, Jarrod. 2012. "Securitization, Social Identity, and Democratic Security: Nixon, India, and the Ties That Bind." *Internaitonal Organization* 66(1):63–93.

Herrmann, Richard K and Jonathan W Keller. 2004. "Beliefs, values, and strategic choice: US leaders' decisions to engage, contain, and use force in an era of globalization." *Journal of Politics* 66(2):557–580.

Herrmann, Richard K, Philip E Tetlock and Matthew N Diascro. 2001. "How Americans think about trade: Reconciling conflicts among money, power, and principles." *International Studies Quarterly* 45(2):191–218.

Herrmann, Richard K., Philip E. Tetlock and Penny S. Visser. 1999. "Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93(3):553–573.

Herrmann, Richard K., Pierangelo Isernia and Paolo Segatti. 2009. "Attachment to the Nation and International Relations: Probing the Dimensions of Identity and their Relationship to War and Peace." *Political Psychology* 30(5):721–754.

Hurwitz, Jon, Mark Peffley and Mitchell A Seligson. 1993. "Foreign Policy Belief Systems in Comparative Perspective." *International Studies Quarterly* 37(3):245–270.

Johns, Robert and Graeme AM Davies. 2012. "Democratic peace or clash of civilizations?" *The Journal of Politics* 74(04):1038–1052.

Johns, Robert and Graeme AM Davies. 2019. "Civilian casualties and public support for military action: Experimental evidence." *Journal of Conflict Resolution* 63(1):251–281.

Kertzer, Joshua D., Deborah Jordan Brooks and Stephen G. Brooks. 2021. "Do Partisan Types Stop at the Water's Edge?" *Journal of Politics* 83(4):1764–1782.

Kertzer, Joshua D, Jonathan Renshon and Keren Yarhi-Milo. 2021. "How Do Observers Assess Resolve?" *British Journal of Political Science* 51(1):308–330.

Kertzer, Joshua D. and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60(1):234–249.

Khong, Yuen Foong. 1992. *Analogies at War.* Princeton, NJ: Princeton University Press.

Krauss, Robert M. and Ezequiel Morsella. 2000. Communication and Conflict. In *The Handbook of Conflict Resolution: Theory and Practice*, ed. M. Deutsch and P.T. Coleman. San Francisco: Jossey-Bass pp. 131–43.

Levendusky, Matthew S. 2018. "Americans, not partisans: Can priming American national identity reduce affective polarization?" *The Journal of Politics* 80(1):59–70.

Lupu, Noam. 2016. *Party brands in crisis: Partisanship, brand dilution, and the breakdown of political parties in Latin America.* Cambridge University Press.

Mintz, Alex and Nehemia Geva. 1993. "Why Don't Democracies Fight Each Other? An Experimental Study." *Journal of Conflict Resolution* 37(3):484–503.

Pop-Eleches, Grigore and Joshua Tucker. 2017. *Communism's Shadow: The Effect of Communist Legacies on Post-Communist Political Attitudes.* Princeton, NJ: Princeton University Press.

Pope, Jeremy C and Jonathan Woon. 2009. "Measuring changes in American party reputations, 1939—2004." *Political Research Quarterly* 62(4):653–661.

Powell, Robert. 2004. "Bargaining and Learning While Fighting." *American Journal of Political Science* 48(2):344–361.

Powers, Kathleen E. 2021. *Nationalisms in International Politics.* Princeton, NJ: Princeton University Press.

Quek, Kai. 2016. "Are Costly Signals More Credible? Evidence of Sender-Receiver Gaps." *Journal of Politics* 78(3):925–940.

Reifler, Jason, Thomas J Scotto and Harold D Clarke. 2011. "Foreign Policy Beliefs in Contemporary Britain: Structure and Relevance." *International Studies Quarterly* 55(1):245–266.

Renshon, Jonathan, Allan Dafoe and Paul Huth. 2018. "Leader influence and reputation formation in world politics." *American Journal of Political Science* 62(2):325–339.

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

Rosch, Eleanor and Barbara Bloom Lloyd, eds. 1978. *Cognition and categorization.* Lawrence Erlbaum Associates, Publishers.

Rousseau, David L. and Rocio Garcia-Retamero. 2007. "Identity, Power, and Threat Perception: A Cross-National Experimental Study." *Journal of Conflict Resolution* 51(5):744–771.

Schultz, Kenneth A. 2001. *Democracy and coercive diplomacy.* New York, NY: Cambridge University Press.

Smith, Alastair and Allan C Stam. 2004. "Bargaining and the Nature of War." *Journal of Conflict Resolution* 48(6):783–813.

Tajfel, Henri and John Turner. 1986. The Social Identity Theory of Intergroup Behavior. In *Psychology of intergroup relations*, ed. S. Worchel and W. Austin. Chicago: Nelson-Hall.

Thies, Cameron G. 2002. "Progress, History and Identity in International Relations Theory: The Case of the Idealist–Realist Debate." *European Journal of International Relations* 8(2):147–185.

Tomz, Michael R. and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107(4):849–865.

Yarhi-Milo, Keren, Joshua D. Kertzer and Jonathan Renshon. 2018. "Tying Hands, Sinking Costs, and Leader Attributes." *Journal of Conflict Resolution* 62(10):2150–2179.

Yilmaz, Emine, Javed Aslam and Stephen Robertson. 2008. A New Rank Correlation Coefficient for Information Retrieval. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* pp. 587–594.