Democratic Credibility Revisited*

Aleksei Opacic† and Lucy Song‡

Harvard University

December 14, 2020

Abstract

The lack of consensus on the existence of a democratic advantage in generating credible commitments in the context of international security results from several theoretical and methodological shortcomings that have characterised the literature to date: inattentiveness to problems of selection bias in observed datasets, disregard for the causal ordering of the relevant variables, and lack of attention to the variations among democracies and forms of audience costs. This paper rejuvenates, and brings more robust evidence to, the long-running debate on the existence of the democratic advantage in international security by establishing the causal effect of a state's regime type on its compellent threat success and credible commitment. Our contribution is threefold. First, we restate existing theoretical propositions in an explicitly causal framework and lay bare the identification assumptions required to make causal inferences about the democratic advantage from observational data. Next, we construct a novel dataset of randomly selected "peacetime dyads" that aims to make more plausible the identification assumptions required to draw causal inferences. Finally, we adapt a series of multiply robust estimators and implement them using novel debiased machine-learning techniques. Our results provide strong support for the democratic credibility hypothesis.

1 Background and Motivation

Among the most productive and influential scholarship in International Relations is the study of the relationship between regime type and international conflict behavior. Existing literature presents

*We would like to thank Katherine Irajpanah, Joshua Kertzer, Andrew O'Donohue, and Katherine Tucker for their comments and feedback.

†Department of Sociology, Harvard University, Cambridge, Massachusetts 02138, USA. Email: aopacic@g.harvard edu.

‡Department of Government, Harvard University, Cambridge, Massachusetts 02138, USA. Email: lucysong@g.harvard.edu.

diverging evidence for the advantage democracies possess in international conflict and crisis bargaining. The prominent if controversial audience costs theory [Tomz, 2007, Trachtenberg, 2012, Fearon, 1994] and the democratic credibility theory [Schultz, 2001] both posit that democracies possess an advantage over non-democracies in their ability to credibly signal diplomatic intentions and threats, owing to their transparent electoral processes that illustrate the direct relationship between the political tenure of the leader and her electorate, which renders it more difficult for the leader to back down after issuing threats on the international stage. By contrast, autocracies, owing to their opaque political procedures, do not often have a credible domestic audience capable of acting as a constraint on the autocrat's international behavior. Therefore, autocratic leaders cannot issue threats and diplomatic statements as credibly as can democratic leaders. The empirical implication of the democratic credibility theory and the associated audience costs theory would indicate that democracies are less likely than autocracies to issue militarized compellent threats and escalate diplomatic crises militarily [Tomz, 2007]. However, democracies are also more likely to win the military conflicts they do initiate, because they tend to start conflicts in which they believe they would prevail in order to avoid sustaining audience costs. Moreover, the threats democracies issue are also more credible to foreign states due to democratic leaders' accountability to their domestic constituents. By contrast, autocracies[1] are less conservative with conflict initiation, but are more likely to lose the military conflicts they initiate. In other words, ceteris paribus, democracy has a positive effect on a state's probability of victory and threat success, and a negative effect on a state's likelihood to initiate militarized conflicts. Democracies' unwillingness to initiate conflict and success in compellence and deterrence threats are often cited as the main reason for the observed democratic peace [De Mesquita, 1999, Fearon, 1994, Schultz, 2001].

However, burgeoning literature in the past decade has challenged earlier theories about such democratic advantage through finding empirical evidence that suggests, contrary to the democratic credibility theory and audience costs theory, autocratic leaders also have mechanisms through which they can achieve similar levels of domestic constraint as democratic leaders. The Autocratic International Relations subfield has thereby sought to challenge the conventional wisdom that democratic regimes, owing to their transparent domestic institutions, possess an advantage over autocracies with regards to crisis bargaining and credible commitments on the international stage. In other words, it is conceivable that there exists "autocratic credibility," and democracies do not possess an advantage in credible commitments and conflict success over certain types of autocracies [Weeks, 2012, Weiss, 2014, Downes and Sechser, 2012].

[1] In this article, we use the terms "autocracies" and "non-democracies" interchangeably.

2 The Democratic Advantage, Challenged

Alexander Downes and Todd Sescher mount a notable challenge to the democratic credibility hypothesis--that democracies are more likely to achieve compellent threat success--with a quantitative analysis of all militarized compellent threats that occurred between 1918 to 2001 [Downes and Sechser, 2012]. They construct a new dataset, Militarized Compellent Threats (MCT), that includes all interstate compellent threats (threats designed to persuade the opponent to change the status quo, where military force is threatened as a punishment for failing to comply with the demand in question) between 1918 - 2001. This dataset provides a more appropriate operationalization of crisis bargaining and audience cost generation during interstate militarized disputes, as it contains interstate conflicts that involved a state issuing threats through either cheap talk (rhetorical threat only) and/or sunk costs (demonstration of military capabilities). Past literature that tests the relationship between regime type and conflict success bases their analysis on the MIDs dataset, which comprises of interstate militarized disputes, most of which do not involve threats. Therefore past scholarship measures the relationship between regime type and success in conflicts that did not involve an explicit audience costs component, a statistical quantity that does not capture the one speicifed by the democratic credibility hypothesis. Downes' and Sescher's findings conclude that democracies do not possess a significant advantage over autocracies in issuing credible compellent threats, as democracies cannot propel their opponent to comply with their demands at a higher rate than autocracies. Their study is exemplary of an emerging body of literature that introduces empirical skepticism to the notion that democracies enjoy special advantages in international crises, at least in the form of credible commitments and audience costs generation.

3 Selection Problems in Previous Studies

However, Downes and Sescher's analysis, as well as other prominent works in Autocratic International Relations, suffers from two related areas of weakness. First, while there have been efforts at disaggregating autocracies based on their institutional [Weeks, 2012, Hyde and Saunders, 2020] and non-institutional attributes [Weiss, 2014], much of existing scholarship still assumes democracies to be monolithic, especially in terms of their non-institutional attributes, when there could be variations in such attributes that affect their willingness and ability to initiate compellent threats. To our knowledge, in the context of the democratic peace debate, there has been little to no attempt at disaggregating democracies by a selection of non-institutional attributes (such as military expenditure

budgets, national conscription policies, female political participation, colonizing history, degree of political polarization) that affect both its propensity to issue threats and likelihood of threat success. Second and relatedly, democracies that select into issuing compellent threats in the first place may differ from democracies that do not across a variety of institutional and non-institutional features. When the chance of gaining concessions from a threat is small, only the most resolute and aggressive types of democracies will make such threats. This further highlights the selection bias of democracies that choose to issue compellent threats. Moreover, democracies may also select on rival states based on threat perception and likelihood of victory (measured through proxy variables such as religion, race, status as former colony, GDP, military power). Without accounting for the selection processes by which democracies decide to initiate threats and militarized disputes, previous attempts to assess the effect of regime type on threat success would yield biased estimates, because they are selecting on a group of democracies that are not representative of democracies in general, and then making a conclusion about the success of democracies in issuing threats at large. Therefore, instead of assuming that all democracies hold similar normative and social values, we measure non-institutional dimensions of democracies across several novel civil society indicators: female political participation, national education, and militarism. We expect that democracies with lower degrees of female political participation and national education but higher higher degrees of militarist culture tend to self select into dispute and threat initiation.

To formalize this selection problem, consider Figure 1, which is a representation of a potential data-generation process non-parametrically in the form of a DAG:

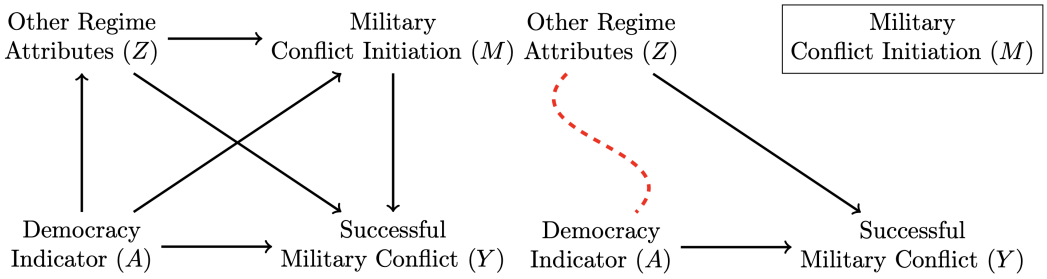

Figure 1: Conditioning on Military Conflict Initiation (a Collider) Induces a Spurious Association between Regime Type and Other Predictors of Conflict Success.

The left panel shows that a vector of variables (X) affect democracy[2], and a further set of variables

[2]For ease of exposition, we do not show confounders of democracy (X), though we assume that these variables (such as contiguity and state B regime status) are predictors of both regime type, institutional attributes, mediator value and compliance.

(Z) such as the variables differentiating democracies described above may affect both the probability of initiating a militarised conflcit and successful military conflict. Militarised conflict, which is often the variable implictly conditioned on by nature of the observed units in IR datasets, is a collider variable - a function of democracy and regime attributes Z (see Elwert and Winship 2014). Downes and Sechser's statistical quantity is the difference in threat success probability among democracies and autocracies who participate in militarized or threat-based activity net of other regime characteristics. However, by confining the population of regime types to those that have engaged in some form of miliarized activity or threat, Downes and Sechser 2012 inadvertently condition on a potential collider variable. In consequence, the authors' empirical quantity only equals their target estimand under implausible identification assumptions--namely that militarized activity is not a consequence of regime type and other regime characteristics that are themselves associated with regime type and which affect the outcome of interest. If these assumptions do not hold, then their subsample will exhibit a spurious dependency between regime type and regime characteristics, and the estimator will be subject to collider stratification bias. The effect of regime type on the sucess of threat may in reality be a function of other regime attributes.

Technically, the data generation process of interest should be conceptualized in a mediation framework. Because mediation invokes counterfactual assignments of both the treatment and the mediator, determinants of both the treatment and mediator need to be adjusted for. To see the implications of the disproportionate selection of democracies into threat-issuing conflicts more formally, let A denote the treatment (an indicator for whether a regime is democratic in a given year), M the mediator (a two-level variable capturing whether or not a state engages in a militarized dispute or threat), and Y the observed outcome (indicating whether a regime was successful - i.e. achieved compliance or victory - in a dispute). Then X and Z are pre-treatment confounders of regime status A and post-treatment confounders of conflict initiation M, respectively.[3] We can consider the conditional expectation function (CEF) $\mathbb{E}[Y|a, X = x]$, the conditional probability of conflict success under regime status a (treating $M = 1$ as fixed):

$$\mathbb{E}[Y|a, X = x] = \int \mathbb{E}[Y|a, X = x, Z = z]dF(Z = z|a, x)$$

There are two ways in which the CEF $\mathbb{E}[Y|a, X = x]$ may be equal to $\mathbb{E}[Y|a', X = x]$. First, it could be the case that, contrary to what is suggested by audience cost theory, democracies genuinely do

[3]It should be noted that Z function as both intermediate confounders and mediators, but given that our focus is on the effect of regime status on Y conditioning only on M, we treat Z exclusively as intermediate confounders of a single mediator M (militarised conflict initiation) rather than as mediators (i.e. rather than exploring the effect of regime status A on Ycontrolling for Z and M jointly).

not initiate conflicts that they are more likely to win, for democracies and autocracies with equivalent values of Z (that is, with the same values of post-treatment confounders):

$$\mathbb{E}[Y|a, X = x, Z = z] = \mathbb{E}[Y|a', X = x, Z = z]$$

However, it could also be the case that:

$$f(z|a, x) \neq f(z|a', x)$$

This latter quantity reflects the distortion in the distribution between A and Z induced by conditioning on the collider variable 'initiated military conflict' - that is, the selectivity of democracies into conflicts. If Z affects probability of conflict success, then an observed equality between democracies and autocracies in their probability of military conflict success may result from the different conditional distributions of Z by regime type. In practice, it is often difficult to predict the direction of collider-stratification bias. However, if we assume the above DGP portrayed in Figure 1 (and Z is negatively associated with Y) it is plausible that, without appropriately accounting for these post-treatment variables (Z), the estimated effect of regime type may be significantly attenuated. To give a concrete illustration: let us assume that democracies engaging in militarized activity are disproportionately paternalistic towards a select number of autocracies due to the autocracies' colonial history and racial and religious composition. This endogenous selection will then open up a non-causal path between A and Y (the probability of military success) if paternalism negatively affects Y (for instance, if paternalism causes a country to overestimate its abilities and include calculations in the payoff model that don't depend just on victory/outcome of the war). The result will be a significantly attenuated estimate of the true effect of A on Y.

4 Effect Heterogeneity by Dyad Type and Audience Cost Type

We introduce further nuance in our estimation of the CDE of regime type on compellent threat success in two innovative ways. First, we specify the regime type pairing of the dyads–democracy-democracy, democracy-autocracy, autocracy-democracy, and autocracy-autocracy. This enables us to determine the extent to which the democratic advantage is a dyadic phenomenon–that is, if democracies possess an advantage in threat success only against autocracies, or, if the democratic advantage is a monadic phenomenon–that is, if democracies possess an advantage in threat success against all regime types. If the dyadic phenomenon is more pronounced than the monadic counterpart (in terms of absolute

value), then perhaps regime familiarity and commonality could be a contributing factor to a state's conflict success. Moreover, analyses of democracy-autocracy and autocracy-democracy dyads allow us to compare the advantage democracies possess in threat success when they are the initiator with the advantage democracies possess in resisting the threat and calling their opponent's bluff when they are the defendant. Second, we filter the MCT dataset by the dichotomous demonstration variable, in order to subset threats that involve demonstrations of military capabilities and threats that involve rhetorics only. This allows us to disaggregate audience costs into its two forms--sunk costs and cheap talk and compare the democratic advantage in compellent threats conditional on the two forms. We expect the democratic advantage to be more pronounced in the latter. Cheap talk is more credible from democracies due to their domestic audience constraint, whereas all regime types can sink costs and resources to demonstrate military might and resolve.

5 Recasting the Democratic Advantage

By stating precisely our theoretical estimand, identification assumptions and estimation processes, we aim to provide a more robust analysis of the effect of regime type on success of military conflict. Our primary theoretical contribution is to recast Downes and Sechser's (2012) test of the democratic credibility and audience costs theory into a causal typology. Specifically, we are interested in establishing the difference in probability of conflict success we would observe by altering the regime status of a state that initiates a militarised threat if, conditioning on regime type, these states were a representative sample of all states[4].

More formally, we draw on the potential outcomes framework (Rubin, 1974; Holland, 1986) to provide model-free definitions of our estimands of interest. Let A, M, Y, X and Z be as above[5]. Then, $Y_i(a, m)$ is the potential outcome of regime i under treatment value a and mediator value m. Our estimand, therefore, is a 'controlled direct effect' (CDE) which captures the average effect of changing a state's regime status from a to a' while fixing the mediator at a given level m. Importantly, note that because M is an indicator capturing whether or not a state engages in a militarised dispute, our outcome is undefined for units that do not initiate a militarised conflict (that is, when $M = 0$). Our theoretical estimand can then be written as:

[4] In Section 6.2 we discuss the population of states that it is most interesting to make claims about.

[5] A denotes the treatment (an indicator for whether a regime is democratic in a given year), M the mediator (a two-level variable capturing whether or not a state engages in a militarized dispute or threat), and Y, the observed outcome (indicating whether a regime was successful - i.e. achieved compliance or victory - in a dispute); X and Z are pre-treatment confounders of regime status A and post-treatment confounders of conflcit initiation M, respectively

$$CDE(a, a', M = 1) = \mathbb{E}[Y(a, M = 1) - Y(a', M = 1)] \tag{1}$$

The CDE is identified under the assumptions of consistency and sequential ignorability Robins, 1997, VanderWeele and Vansteelandt, 2009:

1. consistency: for any unit, if $A = a$ and $M = m$, then $Y = Y(a, m)$;

2. sequential ignorability: $Y(a, m) \perp A|X, \forall a, m$ and $Y(a, m) \perp M|X, A, Z, \forall a, m$.

Sequential ignorability is satisfied in Figure 1, which contains a directed acyclic graph (DAG) summarizing a set of hypothesized causal relationships between the variables outlined previously. Under these assumptions, the CDE can be nonparametrically identified using the following formula:

$$CDE(a, a', M = 1) =$$
$$\int\int \mathbb{E}[Y|A = a, Z = z, X = x, M = 1]dF(z|x, a')dF(x) - \int\int \mathbb{E}[Y|A = a', Z = z, X = x, M = 1]dF(z|x, a')dF(x) \tag{2}$$

Stated simply, by marginalising the conditional probability of conflict success given regime type and post-treatment confounders over the *population* distribution of these confounders we can evaluate the controlled direct effect of regime type on probability of conflict success while purguing for selection processes that could confound our estimates. This estimand can further be adapted to reflect a quantity which we refer to as the 'conditional controlled direct effect' (CCDE), which is the controlled direct effect of regime type on target compliance among observations with a particular set of characteristics. As discussed in Section 4, our primary theoretical interest is in effect heterogeneity by state B regime type. In the next section, we introduce novel non-parametric methods that enable us to estimate this quantity empirically.

6 Estimation

6.1 Methods

To assess the CDE of regime type on conflict success, as discussed previously, it does not suffice to simply regress conflict success on a democracy indicator A and pre-treatment covariates X, and interpret βA_i as a direct effect. This is because it is necessary to adjust for post-treatment variables

Z that confound the assignment of the mediator (conflict initiation, M) to avoid collider stratification bias of the type potentially present in Downes and Sechser 2012. Nevertheless, directly adjusting for Z using standard regression- or matching-based methods will also not identify the CDE. This is because conditioning on Z blocks the causal pathway $A \rightarrow Z \rightarrow Y$ which is part of the CDE, and further unblocks non-causal pathways if Z is a common consequence of A and unobserved determinants of the outcome Y.

Our estimation procedure adapts the multiply robust estimators presented in Zhou 2020, which do not rely on the correct specification of any particular nuisance function. As an illustration of one of the estimators we fit, consider the following functions:[6]

$$\begin{aligned}
\mu_y(x,z) &:= \mathbb{E}[Y|x,a,z,M=1] \\
\nu_y(x) &:= \mathbb{E}_{z|x,a}\mu_y(X,Z) \\
\pi_a(x) &:= Pr[A=a|x] \\
\pi_m(x,z) &:= Pr[M=m|x,a,z]
\end{aligned}$$

In addition, consider the following two-step estimator of $\nu_y(x)$:

$$\hat{\nu}_y(X;\hat{\mu}_y) = \hat{\mathbb{E}}[\hat{\mu}_y(X,Z)|X,A=a]$$

Under the assumption of sequential ignorability described above, $\mu_y(x,z) = \mathbb{E}[Y(a,m)|x,a,z]$ and $\nu_y(x) = \mathbb{E}[Y(a,m)|x]$. Letting $\hat{\mu}_w(x,z)$, $\hat{\nu_y}(x)$, $\hat{\pi_w}(x)$, and $\hat{\pi_w}(x,z)$ denote estimates of these four functions, our primary multiply robust estimator is as follows:

$$\begin{aligned}
&CDE(a,a',M=1) = \\
&\frac{1}{n}\sum_{i=1}^{n}\Bigg[\Big(\hat{\nu}_y(X;\mu_y) + \frac{\mathbb{I}(A=a)}{\hat{\pi}_a(X)}(\hat{\mu}_y(X,Z) - \hat{\nu}_y(X;\hat{\mu}_y)) + \frac{\mathbb{I}(A=a)}{\hat{\pi}_a(X)\hat{\pi}_m(X,Z)}(Y - \hat{\mu}_y(X,Z))_i \\
&\quad - \Big(\hat{\nu}_y(X;\mu_y) + \frac{\mathbb{I}(A=a')}{\hat{\pi}_{a'}(X)}(\hat{\mu}_y(X,Z) - \hat{\nu}_y(X;\hat{\mu}_y)) + \frac{\mathbb{I}(A=a')}{\hat{\pi}_{a'}(X)\hat{\pi}_m(X,Z)}(Y - \hat{\mu}_y(X,Z))_i\Big)\Bigg] \quad (3)
\end{aligned}$$

In addition, we adapt this multiply robust estimator to examine the 'conditional controlled direct

[6]Note that $\mu_y(x,z)$ is the only of the four functions fitted only to observations that initiated a compellent threat, while the other three are fitted over the entire sample of $M=1$ and $M=0$. Our specification therefore differs slightly from that in Zhou 2020, where M is included as a regressor in $\mu_y(x,z)$ and outcomes imputed $M=m$ for all units, since in our case the outcome is undefined for $M=0$.

effect' (CCDE) of regime status on conflict compliance by regime B type[7]. In practice, these nuisance functions can be estimated using flexible machine learning methods and cross-fitting, which has an obvious advantage of reducing model-dependence and thus bias when the true form of the data-generation process is unknown and when non-linearities and interactions between the covariates likely exist. Importantly, the estimating equations for our multiply robust estimators are Neyman orthogonal, and thus estimates produced from machine learning approaches are semiparametric efficient when estimates of the nuisance functions all converge at faster-than-$n^{-1/4}$ rates [Zhou, 2020, Chernozhukov et al., 2018, 2017]. In our main analysis, we use five different machine-learning algorithms to fit the nuisance functions (Gradient Boosting Machine (GBM), Support Vector Machines (SVM), Elastic Net, Lasso Regression and Random Forest), resulting in $5^4 = 625$ estimates of the CDE of regime type on probability of compliance).[8] In our results section, we show only the estimators that employ the same algorithm for all four nuisance functions, but we show the full range of results in Appendix B. We supplement our main analyses with an alternative (semi-)parametric estimation procedure (a pure imputation estimator), whose results we explore in Appendix B also.

6.2 Data

Confronting the "democratic selection problem" has the obvious difficulty in that many of the datasets commonly used in International Relations implicitly condition on a post-treatment variable, such as military conflict (a form of sample truction bias). As such, it is not possible to estimate the conditional distribution $E[Z|x, a]$, which describes the distribution of intermediate confounders marginalised over M, and therefore not possible to employ the (semi-)parametric methods described above. This motivates our need to find a suitable control group of states for which $M = 0$ - that is, those states which do not issue a compellent threat. However, if we were to establish a control group based on country year observations for every M = 0 state–the states that did not engage in a conflict and did not initiate a compellent threat–the control group would be very large compared to the treatment group, as the international system is relatively peaceful most of the time. The high ratio of treatment group observations to control group observations is a common conundrum in empirical International Relations, where many events of interest ($Y = 1$s) are rare.

However, many of the dyads in the control group have few interactions with each other (e.g. Botswana and Grenada), and thus are "nearly irrelevant" [Maoz and Russett, 1993, King and Zeng,

[7] Our methodological contribution in this piece is therefore to employ Zhou's 2020 triply robust estimator for conditional controlled direct effects. In practice, we can estimate this quantity via the conditional expectation of the signals returned by the estimator [see Semenova and Chernozhukov, 2020].

[8] We use the empirical analog of the efficient influence function as shown in Equation 3 to construct standard errors.

2001 in the sense that they provide little information on our quantity of interest–the effect of regime type on the credibility and success of compellent threats, among a population of dyads that have at least some degree of interaction with each other. To resolve this problem, we adopt a more efficient sampling and control group design. We construct our control group in a 'case-control' style with a random stratified sample of the dyads in the control group. Random selection is desirable because the selection rule is known to be independent of all other variables (as long as the sample size is large enough) and therefore cannot cause bias [King and Zeng, 2001]. In order to ensure that the dyads in our control group are relevant and interactive (e.g. the US and Japan instead of the US and Madagascar), we sampled from the complete population of peaceful dyads (where $Y = 0$) with an alliance score higher or equal to the median alliance score of the given year. Moreover, to account for the positive correlation between contiguity and conflict, our control group consists of an even split between randomly-selected non-contiguous and contiguous dyads. In order to compensate for differences in the sample ($\bar{m}$) and population (τ) fractions of 'ones' (i.e. states observed in conflict) induced by choice-based sampling, we weight the data following the procedure suggested in King and Zeng 2001.[9]

In sum, our treatment group consists of states that have issued a compellent threat and engaged in a militarized dispute (when M=1), which we obtained using the MCT dataset. Our control group consists of a randomly sampled set of state dyads that were pacific in a given year. Our estimand of interest is therefore the CDE defined in Equation 1, which, specifically, captures the effect of both changing a state's regime status from non-democratic to democratic *and* intervening to make the state initiate a compellent militarised threat, averaged over *all dyads of with at least some degree of regular interaction with each other,* as measured by contiguity or alliance score. As is customary, we measure democracy using the revised combined polity score (POLITY2) in the latest version of the Polity dataset, PolityIV. We also created a dicohtomas democracy variable, with states with a POLITY2 score of 6 or higher classified as democracies, per convention.

In addition to using the control variables specified in Downes' model, we have also constructed a set of post-treatment variables (Z) in order to account for regime attributes other than regime type (A) that may affect the state's probability of initating a compellent threat and conflict and probability of victory. We have constructed four post-treatment indicators: gender political representation, regional power status, aggregate levels of education, and military expenditure quintile. The aggregate gender indicator consists of the year women's suffrage was granted and the share of female politicians in

[9]The weights are given by $w_i = w_1 M_i + w_0(1 - M_i)$, where $w_1 = \frac{\tau}{\bar{m}}$ and $w_0 = \frac{1-\tau)}{(1-\bar{m})}$, and where τ is the population proportion of 1s and $\bar{m}$ is the sample proportion of states observed with M=1.

parliament. Bourgeoning research on gender and international relations posits that, for a variety of biological and enviornmental reasons, states with more women's representation at populist and leadership levels tend to be more pacific than states with lower degrees of gender equality in political representation Barnhart et al. 2020. The world and regional power status indicators rank states into quintiles by their composite index of national capability (cinc) scores, within their respective regions and compared to the rest of the world. The military expenditure quintile is similar, and measures a state's annual military budget relative to all other states. We also include a binary indicator variable for the existence of conscription (mandatory military service policies) in a state, which may be a function of a state's militarism. We expect states with higher degrees of militarism to be more likely to initiate compellent threats. We also include a national education variable, as states with higher national levels of education could be less conflict-prone (education and inequality? education and progressive politics?) The Z variables also enable us to distinguish conceptually between variables that are pretreatment and post-treatment, with the latter being a function of regime type.

Since some components of X and Z contain a number of missing values, we impute missing values via multiple imputation and adjust the standard errors of our parameter estimates using Rubin's 2004 method.

7 Results

7.1 Replication of Downes and Sechser (2012)

We begin by replicating and extending Downes and Sechser 2012 (henceforth DS), by both improving the presentation of the results (which are presented only in log odds form with standard errors in the original piece) and attending to potential model misspecification. Figure 2 presents estimates from DS' original model specifications. We provide predicted probability point estimates and 95% confidence intervals where the standard error of the estimate is derived from a simulation of the coefficients from the multivariate normal distribution. Consistent with D&S, the Figure provides strong evidence for no effect of democracies on compliance of the target.

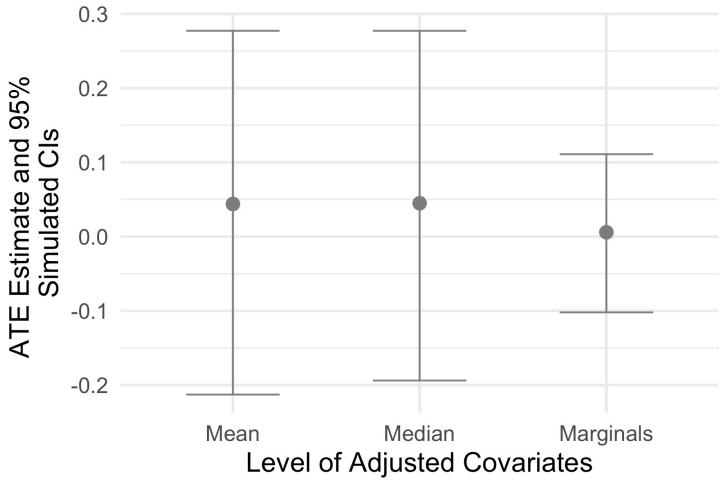

Figure 2: Predicted probabilities from Downes and Sechser (2012).

Further analyses (shown in Appendix C) show that no observations lie inside the convex hull and further, that predictions of both counterfactual and factual units are considerably model dependent. This motivations the following procedure, in which we use Chernozhukov et al's 2018, 2017 double/debiased machine-learning estimator to re-analyse the DS subsample. We use random forests and support vector machines to estimate the propensity score and outcome models and estimate standard errors from the variance of the emprirical analog of the efficient influence function. This results in $2 \cdot 2 = 4$ estimates of the ATE. The results show that, despite severe model misspecifiction, the use of a fully-flexible non-parametric machine learning approach does not change our estimates of the ATE. We therefore turn to our new dataset to examine the extent of selection bias into this MCT subsample by estimating the CDE of democracy on compliance.

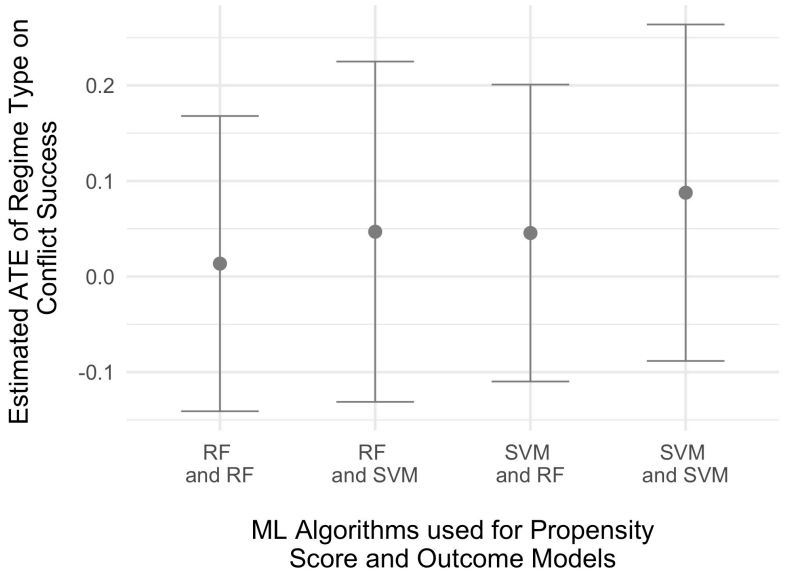

Figure 3: Four machine learning estimators of the ATE fitted to the original DS dataset.

7.2 Selection and democracies

As discussed earlier, conditioning on military conflict initiation will distort the associations between regime type and the covariates compaed with their levels in the general population. As we can see from Figure 4, certain civil social attributes differ insubstantially between autocracies and democracies that initiate a militarised threat, and their general population counterparts. However, the disributions of certain variables are indicative of disproportionate selection of democracies into initiating a threat.

First, we note several general differences between autocracies and democracies across civil social attributes. In general, democracieshave a higher average level of education, female suffrage, conscription, and military capabilities. Specifically, we observe a significant difference between the average level of conscription between the regime types that select into initiating compellent threats. A higher percentage autocracies and democracies that initiate compellent threats have conscription policies compared to their pacific counterparts; this is especially true of autocracies that initiate militarized threats, whsoe conscription rates are roughly half of those of pacific autocracies. In terms of intra-regime type differences, democracies that initiate compellent threats are on average ranked higher on military quintiles than those that do not. Surprisingly, democracies that initiate compellent threats have a higher average level of education compared to the ones who do not. Moreover, democracies that initiate threats do not seem to differ from democracies that do not in terms of the level of female suffrage, which provides preiliminary evidence against the suffragist peace [Barnhart et al., 2020]. We

also note that the measure for female suffrage in autocracies does not carry much significance, as many autocracies do not have fair and free elections.

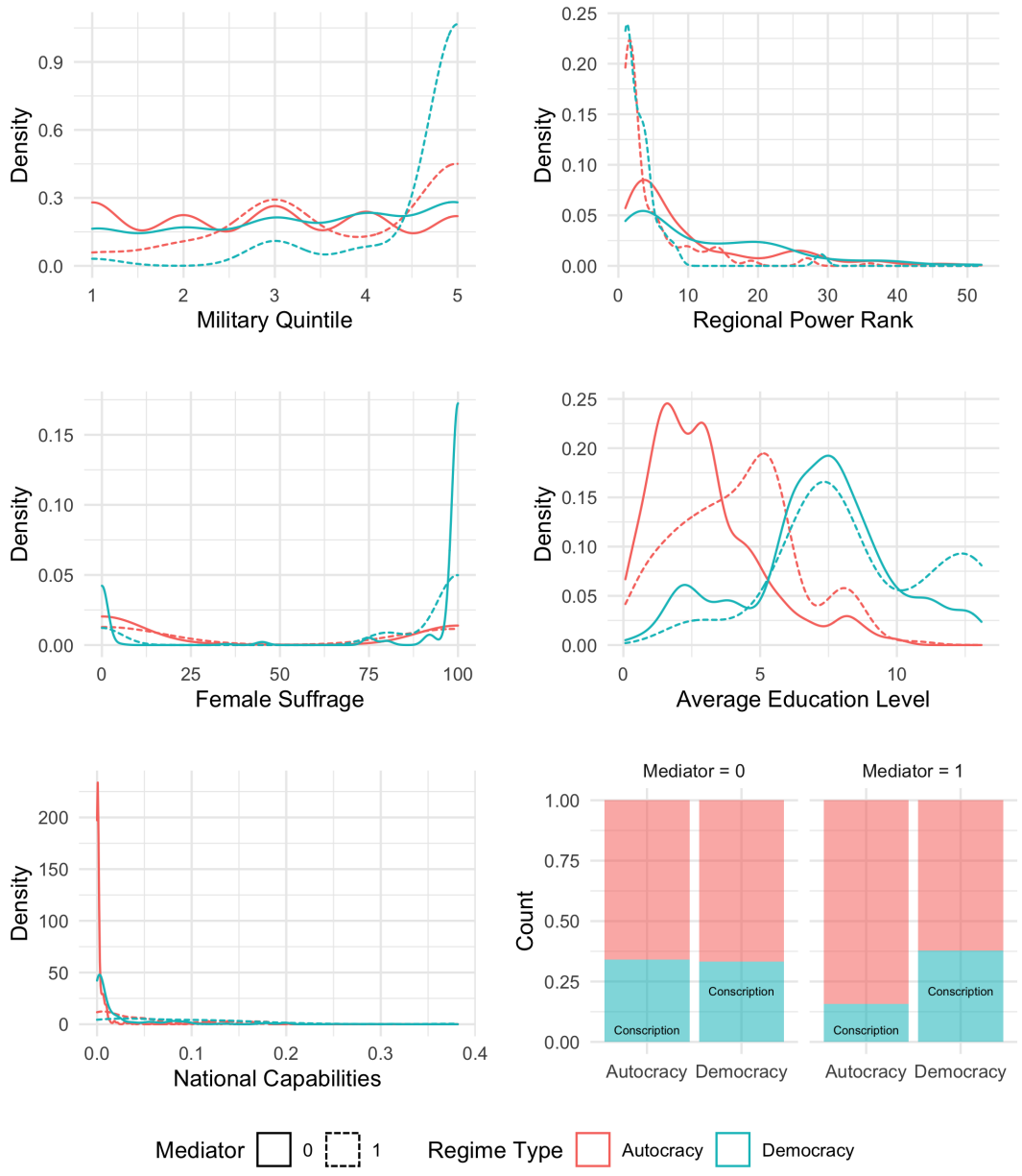

Figure 4: Covariate Balance Plot.

As discussed in Section 3 above, without accounting for the selection processes by which democracies decide to initiate threats and militarized disputes, efforts at assessing the effect of regime type on threat success would yield biased estimates due to the distortion between these covariates and regime status. In order to overcome this problem of "collider bias," we implement the multiply robust

estimators described above. We first use five machine-learning algorithms (GBM, SVM, Elastic Net, Lasso Regression and Random Forest) to fit the nuisance functions, and then use our triply robust estimator shown in Equation 3 to construct our multiply robust estimators of the CDE of regime type on probability of conflict success, adjusting for selection processes that would confound estimates made using standard regression-based estimators. Figure 5 plots the controlled direct effects obtained from these five machine-learning algorithms used to fit all of the four nuisance functions. The figure shows gives strong support for the fact that democacies confer a 2- to 6-point advantage on the probability of conflict compliance relative to non-democracies. This provides evidence in favor of the democratic credibility theory, which posits that democracies, owing to their ability to generate audience costs more effectively and their transparent domestic accountability mechanisms, are able to more credibly issue compellent threats and prompt their opponent state to back down and comply with their demands. In other words, regime types differ in their abilities to signal credible commitment during international security crisis bargaining.

This set of findings contrasts strongly with those displayed by Figure 2, where the estimated democratic advantage is statistically and substantially insignificant: all point estimates are distinguishable from zero. This finding is robust to a number of alternative parametric and semi-parametric specifications, as well as across all 625 machine-learning models that we estimate (see Appendices A and B for further details).

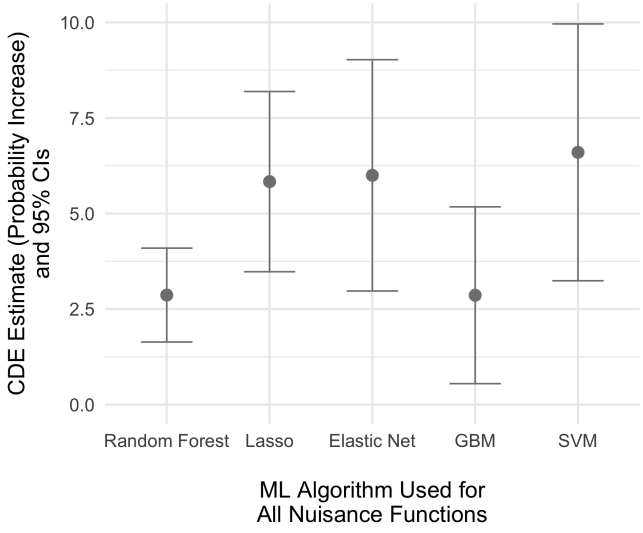

Figure 5: Multiply robust estimates of the controlled direct effect of democracy on conflict complicance with five ML algorithms (GBM, SVM, Elastic Net, Lasso Regression and Random Forest) used to fit all nuisance functions. Error ranges correspond to 95% Wald confidence intervals. Estimates are obtained from 10-fold imputation using random forests; standard errors are derived from the sample variance of the estimated EIF and adjusted for multiple imputation using Rubin's (1987) imputation method..

7.3 Does 'state B' regime type make a difference?

Having established strong evidence in support of a democratic advantage in threat success, we now move on to our final analysis - specifically, an exploration of the 'conditional controlled direct effect' (CCDE) of regime type on conflict compliance by the regime status of state B. As discussed above in Section 4, it is plausible that the democratic advantage is a dyadic phenomenon rather than monadic phenomenon (that is, if democracies possess an advantage in threat success only against autocracies, or, instead, in threat success against all regime types). As expected, democracies enjoy an even higher rate of compellent threat success against autocractic opponents. Specficially, the democratic advantage in threat success against autocracies is in the region of 2-3 times larger than the democratic advantage in threat success against democracies. This provides circumstantial evidence that the dyadic effect is stronger than the monadic effect, and regime similarity may mitigate the credibility of costly signals.

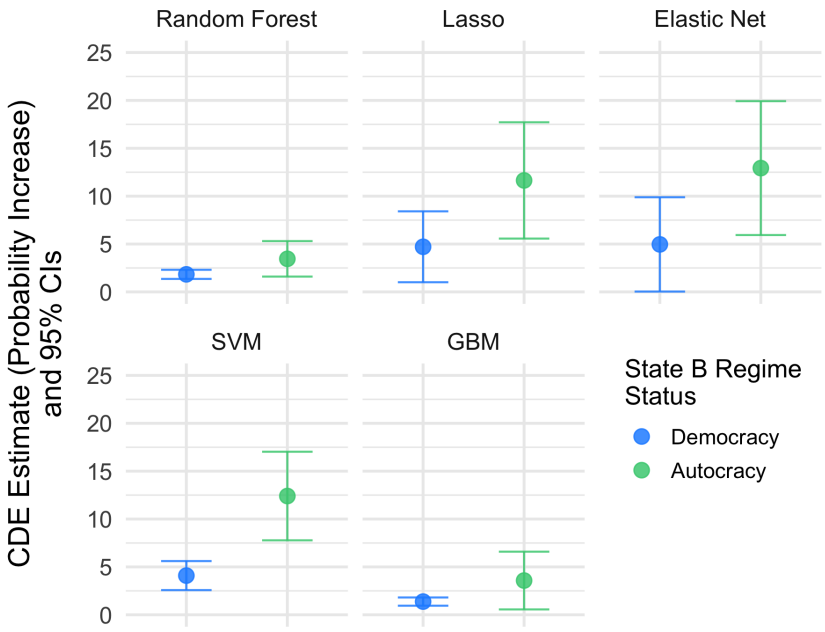

Figure 6: Multiply robust estimates of the conditional controlled direct effect of democracy (given state B regime status) on conflict complicance. Each pair of estimates corresponds to a different ML algorithms used to fit all nuisance functions. Error ranges correspond to 95% Wald confidence intervals. Estimates are obtained from 10-fold imputation using random forests, and standard errors are as in Figure 5.

8 Conclusion

The democratic peace theory is arguably the most significant contribution in the field of International Relations, both in terms of the rigor of its empirical and theoretical findings, as well as the importance of its policy implications. It is therefore crucial that scholars obtain unbiased and robust estimates for the advantage democracies possess in crisis bargaining during international disputes, in order to more accurately ascertain the sources of the democratic peace. Our starting point in this article has been the counterintuitive finding presented in the International Relations literature that there exists no democratic advantage in threat compliance and crisis bargaining in international security Downes and Sechser, 2012, and the possibility of a hitherto unacknowledged form of selection bias that may have plagued earlier results. By adjusting for selection problems using a new dataset and novel machine-learning techniques, and renewing emphasis on variations in civil social attributes of regime types, we have been able to confront the question of a democratic advantage in conflict compliance with greater attentiveness to problems of causal inference than has previously been displayed in scholarly attempts to address this question. In sum, our article offers strong support for the existence of a democratic advantage in international security crisis bargaining. Specifically, democacies confer a 2- to 6-point advantage on the probability of conflict confliance relative to non-democracies, a result that is highly robust over an incredibly large number of model specifications and alternative parameterizations. Further, this advantage is far more pronounced when the target state is an autocracy: the democratic advantage in threat success against autocracies is in the region of 2-3 times larger than the democratic advantage in threat success against democracies. Our article also paves the foundation for future extensions and applications on related topics. In addition to measuring the causal effect of regime type on compellent threat success, it could be worthwhile to measure such effect on deterrant threat success and on conflict success at large. Moreover, it may also be promising to investigate the existence of a democratic advantage in credible commitment and signaling in other areas of foreign policy, such as trade agreements and alliance formations. Overall, our empirical findings highlight the immense potential of applying statistically-rigoroous methods to adjudicate the relationship between regime type and state behavior in international relations.

Appendix

A Additional Machine Learning Estimation

In the main analysis, we display four different estimates of the average treatment effect of democracy on conflict success, as well as two of the conditional average treatment effect given state B's regime type. In practice, since we fit 5 different algorithms for the nuisance functions, this gives us $5^4 = 625$ estimates of our primary estimands, all of which use a differnet algorithm to learn the functional form of the nuisance function from the data. For ease of exposition, we show only a select few here (though the full range of models can be seen via our code in `04_analysis.R`. We can see that our main results are highly consistent under alternative parametrisations of the nuisance funtions. The democratic advantage (controlled direct effect) is estimated as being in the region of 2 to 10 points increase on the probability scale of success (Table 1), and the democratic advantage in threat success against autocracies is in the region of 2-3 times larger than the democratic advantage in threat success against democracies (Table 2).

	Treatment Model	Mediator Model	Outcome Model	Predicted Model	Estimate	SE
1	GBM	GBM	RF	SVM	2.88	1.02
2	RF	RF	SVM	RF	2.88	0.68
3	E-Net	GBM	SVM	SVM	2.11	0.52
4	Ridge	SVM	SVM	RF	7.82	1.34
5	E-Net	GBM	RF	RF	2.08	0.45
6	E-Net	GBM	RF	SVM	2.13	0.45
7	SVM	SVM	SVM	RF	9.49	1.52
8	Lasso	Lasso	SVM	GBM	9.23	2.09
9	GBM	RF	RF	GBM	2.25	0.59
10	Lasso	GBM	SVM	RF	2.10	0.50
11	Ridge	Lasso	GBM	GBM	6.34	1.70
12	SVM	Lasso	GBM	GBM	7.82	2.00
13	Ridge	SVM	SVM	GBM	7.82	1.34
14	Ridge	Lasso	GBM	SVM	6.28	1.70
15	Ridge	SVM	GBM	RF	5.06	0.97
16	SVM	SVM	RF	GBM	9.55	1.57
17	Lasso	RF	RF	SVM	1.89	0.46
18	RF	GBM	GBM	SVM	2.05	0.45
19	Ridge	GBM	GBM	GBM	1.54	0.31
20	Ridge	GBM	SVM	GBM	2.25	0.59

Table 1: 20 Randomly-selected ML estimates of the CDE of democracy on conflict complicance with five ML algorithms.

	Regime B Status	Treatment Model	Mediator Model	Outcome Model	Predicted Model	Estimate	SE
1	Autocracy	ENet	Lasso	GBM	GBM	6.27	1.44
2	Democracy	ENet	Lasso	GBM	GBM	5.08	2.55
3	Autocracy	ENet	RF	RF	SVM	2.30	0.69
4	Democracy	ENet	RF	RF	SVM	1.13	0.08
5	Autocracy	ENet	SVM	GBM	SVM	5.50	1.32
6	Democracy	ENet	SVM	GBM	SVM	3.44	0.89
7	Autocracy	GBM	ENet	SVM	RF	17.37	6.44
8	Democracy	GBM	ENet	SVM	RF	6.09	3.01
9	Autocracy	Lasso	ENet	SVM	GBM	12.58	3.32
10	Democracy	Lasso	ENet	SVM	GBM	5.22	2.66
11	Autocracy	Lasso	GBM	GBM	RF	1.56	0.36
12	Democracy	Lasso	GBM	GBM	RF	1.20	0.20
13	Autocracy	Lasso	RF	SVM	GBM	2.18	0.62
14	Democracy	Lasso	RF	SVM	GBM	1.18	0.09
15	Autocracy	Lasso	RF	SVM	RF	2.18	0.62
16	Democracy	Lasso	RF	SVM	RF	1.18	0.09
17	Autocracy	Lasso	SVM	RF	GBM	9.46	1.82
18	Democracy	Lasso	SVM	RF	GBM	3.14	0.65
19	Autocracy	RF	ENet	GBM	RF	7.98	2.27
20	Democracy	RF	ENet	GBM	RF	7.98	3.96
21	Autocracy	RF	GBM	RF	GBM	4.08	1.22
22	Democracy	RF	GBM	RF	GBM	2.00	0.27
23	Autocracy	Ridge	ENet	GBM	SVM	6.35	1.64
24	Democracy	Ridge	ENet	GBM	SVM	7.11	5.72
25	Autocracy	Ridge	Lasso	RF	GBM	11.91	3.02
26	Democracy	Ridge	Lasso	RF	GBM	6.08	3.98
27	Autocracy	Ridge	Lasso	SVM	RF	11.85	2.91
28	Democracy	Ridge	Lasso	SVM	RF	6.44	4.51
29	Autocracy	Ridge	Lasso	SVM	SVM	11.84	2.91
30	Democracy	Ridge	Lasso	SVM	SVM	6.43	4.50
31	Autocracy	SVM	ENet	GBM	SVM	8.30	2.70
32	Democracy	SVM	ENet	GBM	SVM	6.71	3.36
33	Autocracy	SVM	Lasso	SVM	SVM	14.44	3.42
34	Democracy	SVM	Lasso	SVM	SVM	5.86	1.77
35	Autocracy	SVM	RF	GBM	RF	2.03	0.76
36	Democracy	SVM	RF	GBM	RF	1.86	0.28
37	Autocracy	SVM	RF	GBM	SVM	1.93	0.76
38	Democracy	SVM	RF	GBM	SVM	1.87	0.29
39	Autocracy	SVM	RF	SVM	RF	2.87	0.71
40	Democracy	SVM	RF	SVM	RF	1.72	0.20

Table 2: 20 Randomly-selected ML estimates of the conditional controlled direct effect of democracy (given state B regime status).

B Alternative Estimation Procedures

B.1 Pure Imputation Estimator

In the main analyses, we employ the standard binary Polity definition of democracy [Jaggers and Gurr, 1995], which defines democracies as states with polity scores of +6 or higher (on a scale from −10 to +10), and update the MCT dataset using the latest Polity IV dataset. The use of a binary treatment variable is necessary in order to fit our multiply robust estimators, but in the following we employ a pure imputation estimator for a continous measure of democracy (the unaltered Polity IV Score). Our substantive conclusions are highly robust to this pure-imputation estimator with a continuous democracy measure. Figure 7 displays the predicted probabiltiies of conflict compliance by polity score. The figure clearly shows an increase in point estimate of the predicted probability of compliance from 0.63 to 0.66, which is consistent with the point estimates obtained in our main analysis. The high degree of uncertainty associated with our estimates is a function of few observations per polity score.

Specifically, our estimation procedure is as follows. The estimator can be written as:

$$\mathbb{E}_{Z|X,A}\mathbb{E}[Y|X,Z,A,M=1]$$

To implement this estimator, we: (1) fit a model for the conditional mean of Y given X, Z and A for all units where $M=1$, thus obtaining $\hat{\mathbb{E}}[Y|X,Z,A,M=1]$ (2) Fit a model for the the conditional mean of $\hat{\mathbb{E}}[Y|X,Z,A,M=1]$ given X and A. To evaluate $\mathbb{E}[Y|X,Z,A,M=1]$, we use random forests to estimate the outcome model $\mathbb{E}[Y|X,Z,A,M=1]$ and the propensity score model $P(M=1|X,Z)$, and combining the two results in a doubly robust and debiased estimator which is $\sqrt{n}$-consistent. We then use a parametric model generalised linear model with natural splines for the continuous democracy treatment to estimate the controlled direct effect as outlined in Section 3.

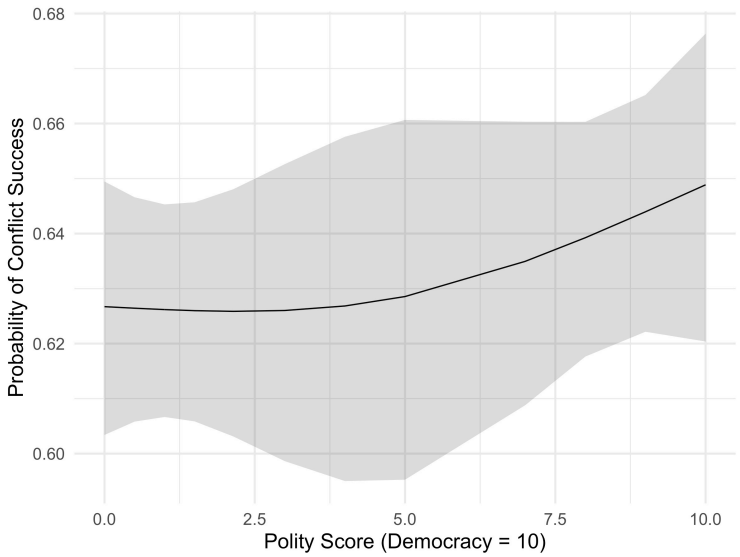

Figure 7: Pure imputation estimator with parametric smoothing for the probability of target compliance by regime type

C Convex Hull Analysis

In supplementary analyses, we find that no observations in the original DS datset lie inside the Convex Hull and that, further, the conclusions from the original DS subsample are highly model-dependent. To show this, we follow the procedure suggested by King and Langche 2006 and estimate predicted factuals and counterfactuals. Specifically, we use (a) the original DS logit specification with no interaction terms as well as (b) a machine-learning doubly robust model employing random forests to estimate the propensity score and outcome models for the factual as well as counterfactual predictions. The following figures show that both counterfactual as well as factual predictions are extremeley model dependent. Nevertheless, in spite of this finding (as we bring out in the main text), DS' original conclusions that there is no democratic advantage in conflict compliance are robust to a range of highly non-parametric specifications. It is the robustness of DS' original results to multiple specifications that motivates our primary analysis to explore potential selection bias of the observed democracies in the DS dataset.

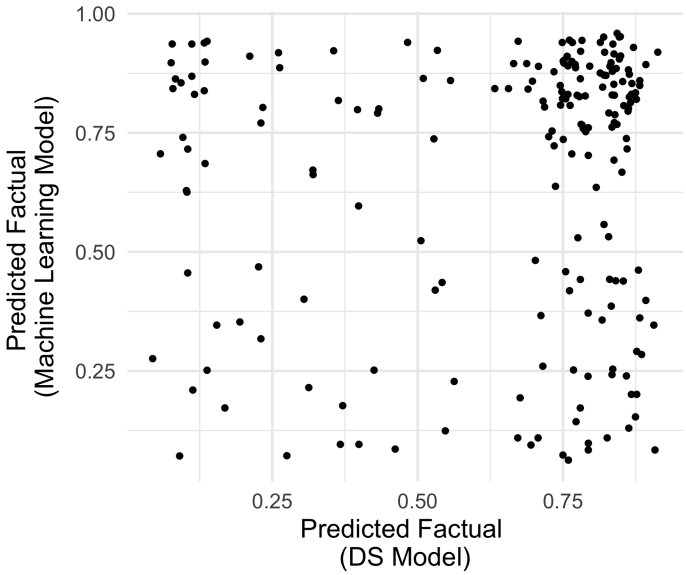

Figure 8: Predicted factuals from DS' original logit specification with no interaction terms as well as a machine-learning doubly robust model employing random forests to estimate the propensity score and outcome models

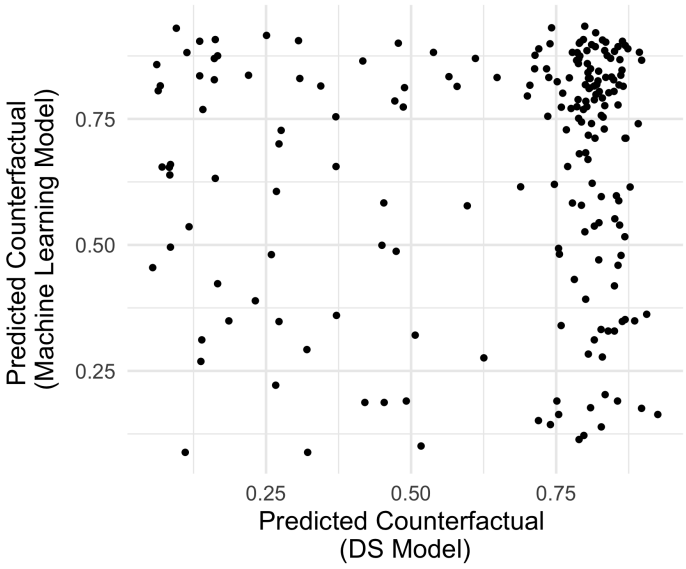

Figure 9: Predicted Counterfactuals from DS' original logit specification with no interaction terms as well as a machine-learning doubly robust model employing random forests to estimate the propensity score and outcome models

References

Joslyn N. Barnhart, Robert F. Trager, Elizabeth N. Saunders, and Allan Dafoe. The Suffragist Peace. *International Organization*, sep 2020. ISSN 15315088. doi: 10.1017/S0020818320000508.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–65, 2017.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, 2018. ISSN 1368423X. doi: 10.1111/ectj.12097.

Bueno De Mesquita. An insitutional explanation of the democratic peace. *Dec*, 93(4):791, 1999.

Alexander B. Downes and Todd S. Sechser. The Illusion of democratic credibility. *International Organization*, 66(3):457–489, jun 2012. ISSN 00208183. doi: 10.1017/S0020818312000161.

James D Fearon. Domestic Political Audiences and the Escalation of International Disputes. *Source: The American Political Science Review*, 88(3):577–592, 1994.

Paul W Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960, dec 1986. ISSN 01621459. doi: 10.2307/2289064. URL http://www.jstor.org/stable/2289064.

Susan D. Hyde and Elizabeth N. Saunders. Recapturing Regime Type in International Relations: Leaders, Institutions, and Agency Space. *International Organization*, 74(2):363–395, mar 2020. ISSN 15315088. doi: 10.1017/S0020818319000365.

Keith Jaggers and Ted Robert Gurr. Tracking democracy's third wave with the polity iii data. *Journal of peace research*, 32(4):469–482, 1995.

Gary King and Langche Zeng. Logistic Regression in Rare Events Data. *Political Analysis*, 9(2): 137–163, 2001. ISSN 1047-1987. doi: 10.1093/oxfordjournals.pan.a004868.

Gary King and Langche Zeng. The dangers of extreme counterfactuals. *Political Analysis*, 14(2): 131–159, 2006.

Zeev Maoz and Bruce Russett. Normative and Structural Causes of Democratic Peace, 1946-1986. *Source: The American Political Science Review*, 87(3):624–638, 1993.

James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. ISSN 1939-2176(Electronic),0022-0663(Print). doi: 10.1037/h0037350.

Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons, 2004. ISBN 0471655740.

Kenneth Schultz. *Democracy and Coercive Diplomacy*. Cambridge University Press, Cambridge, UK, 2001.

Vira Semenova and Victor Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 2020.

Michael Tomz. Domestic audience costs in international relations: An experimental approach. *International Organization*, 61(4):821–840, oct 2007. ISSN 00208183. doi: 10.1017/S0020818307070282.

Marc Trachtenberg. Audience costs: An historical analysis. *Security Studies*, 21(1):3–42, jan 2012. ISSN 09636412. doi: 10.1080/09636412.2012.650590.

Tyler J VanderWeele and Stijn Vansteelandt. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2(4):457–468, 2009. ISSN 1938-7997.

Jessica L. Weeks. Strongmen and straw men: Authoritarian regimes and the initiation of international conflict. *American Political Science Review*, 106(2):326–347, may 2012. ISSN 00030554. doi: 10.1017/S0003055412000111.

Jessica Chen Weiss. *Powerful Patriots: Nationalist Protest in China's Foreign Relations*. Oxford University Press, Oxford, UK, 2014.

Xiang Zhou. Some Doubly and Multiply Robust Estimators of Controlled Direct Effects. *arXiv preprint arXiv:2011.09569*, 2020.