A Framework for Measuring Leaders' Willingness to Use Force: Appendix

Jeff Carter
Department of Government and Justice Studies
Appalachian State University
jtc475@gmail.com

Charles E. Smith, Jr.
Department of Political Science
The University of Mississippi
pssmith@olemiss.edu

April 20, 2020

This appendix contains three sections. The first provides summary statistics for the data underlying our measures, and technical descriptions, diagnostic results, and summary statistics for each of our four models and measures. The second section reports the results tables associated with our validation analyses. The third section describes a set of alternative latent variable models we estimated and explains our preference for the models reported in the main manuscript. Replication materials are available at the American Political Science Review Dataverse: `https://doi.org/10.7910/DVN/7WFX1K`.

1 Information about Data, Models, and Measures

This section contains the following for each of our four models: its technical definition, a set of diagnostic statistics, summary statistics about the resulting measure, and item characteristic curves (ICCs) that jointly characterize how the difficulty (α) and discrimination (β) parameters associated with each background characteristic are related to the estimated latent measure of leaders' willingness to use force (θ).

A Underlying Data

Table A-1 presents a set of summary statistics for the publicly available data from Horowitz, Stam and Ellis (2015), Keller (2005), Seki and Williams (2014), and Brambor, Lindvall and Stjernquist (2017) that our measures are based on.

Table A-1: Summary Statistics for Underlying Data

	N	Mean	St. Dev.	Min	Max
Military Service	2,924	0.308	0.462	0.000	1.000
Non-Combat	2,914	0.086	0.280	0.000	1.000
Combat	2,914	0.220	0.414	0.000	1.000
Win War	2,931	0.076	0.265	0.000	1.000
Lose War	2,928	0.063	0.243	0.000	1.000
Military Career	2,965	0.206	0.405	0	1
Military Education	2,898	0.169	0.375	0.000	1.000
Rebel	2,921	0.262	0.440	0.000	1.000
Rebel Win	2,928	0.044	0.205	0.000	1.000
Rebel Loss	2,928	0.023	0.148	0.000	1.000
Irregular Entry	2,847	0.192	0.394	0.000	1.000
Male Leader	2,965	0.986	0.118	0	1
Older Leader	2,918	0.477	0.500	0.000	1.000
Education Level	2,802	0.450	0.498	0.000	1.000
Spouses	2,360	0.842	0.364	0.000	1.000
Married	2,497	0.056	0.230	0.000	1.000
Married in Power	2,451	0.097	0.296	0.000	1.000
Divorced	2,367	0.899	0.301	0.000	1.000
Number of Children	1,775	0.621	0.485	0.000	1.000
Parental Status	2,342	0.045	0.208	0.000	1.000
Illegitimate	2,965	0.981	0.137	0	1
Royalty	2,965	0.927	0.260	0	1
Orphan	2,965	0.980	0.141	0	1
Teacher	2,965	0.122	0.327	0	1
Journalism	2,965	0.061	0.239	0	1
Law	2,965	0.714	0.452	0	1
Medicine	2,965	0.965	0.185	0	1
Religion	2,965	0.018	0.133	0	1
Activist	2,965	0.113	0.316	0	1
Career Politician	2,965	0.705	0.456	0	1
Creative	2,965	0.056	0.229	0	1
Business	2,290	0.880	0.325	0.000	1.000
Aristocrat/Landowner	2,965	0.929	0.257	0	1
Police	2,965	0.990	0.102	0	1
Science/Engineer	2,965	0.042	0.202	0	1
Blue Collar	2,290	0.083	0.277	0.000	1.000
Keller Index	42	49.372	6.592	33.144	65.646
Hawk	398	0.668	3.490	−17.777	25.000
Right-Left	398	0.593	19.390	−44.500	64.100
International Peace	398	1.892	2.706	0.000	28.261
Ideology	1,199	1.282	0.737	0.000	2.000

We note here our models required us to transform some of the variables from the LEAD project in up to two ways. First, the IRT models we use require that all items/manifest variables be dichotomous. We therefore converted non-dichotomous variables that identified a leader's age, education level, parental status, number of spouses, and number of children to dummy variables coded "1" if the value of the variable was greater than the sample mean and "0" otherwise. Second, the manifest variables in IRT models are assumed to be coded such that a value of "1" implies that an individual scores higher on the latent trait being estimated. In the case of educational testing research, a correct answer on a test implies that a student has more latent ability. In our case, this means that variables must be coded in a way the existence of a given background experience or trait corresponds to a greater latent willingness to use force. There are a number of background experiences and attributes coded as part of the LEAD project that do not clearly imply that a leader should be relatively more or less hawkish given that they have this characteristic (see Ch. 1 of Horowitz, Stam and Ellis (2015) for an extended discussion of this point). We therefore calculated a set of bivariate correlations between the LEAD variables and the initiation of a militarized interstate dispute and recoded dichotomous variables that had a negative correlation such that the absence of a given characteristic would result in a variable taking on a value of "1" and the presence of the characteristic would take on a value of "0." We did this for the following variables: law, medicine, career politician, business, aristocrat or landowner, police, total spouses, married, married in power, divorced, number of children, whether a leader was an illegitimate child, whether a leader was a member of the royalty, and whether a leader was an orphan.

B Model 1

Our first model (M1) is fully characterized in the main manuscript.

B.1 Diagnostic and Summary Statistics

The following tables report diagnostic and summary statistics associated with α_j and β_j, summary statistics for θ_i and a set of item characteristic curves.

Table A-2: α_j Diagnostics for Model 1

	mean	sd	2.5%	97.5%	$\hat{R}$
Military Service	0.940	0.044	0.852	1.027	1.000
No Combat	1.464	0.050	1.364	1.559	0.999
Combat	1.202	0.048	1.108	1.299	1.000
War Win	1.827	0.055	1.721	1.935	1.000
War Loss	2.840	0.078	2.693	2.997	0.999
Military Career	3.059	0.084	2.901	3.229	1.000
Military Education	3.540	0.174	3.298	3.967	1.000
Rebel	4.543	0.488	4.002	5.869	1.001
Rebel Win	1.558	0.050	1.459	1.656	0.999
Rebel Loss	2.777	0.142	2.580	3.140	1.001
Irregular Entry	1.665	0.055	1.558	1.774	0.999

As described in the manuscript, our model assumes $\beta \sim \textbf{Beta}\ (\frac{1}{2}, \frac{1}{2})$, which is the Jeffreys priors over a Bernoulli distribution. A little-known feature of this prior is that it is not defined on Euclidean space, but rather on a Riemannian metric with extremely high curvature in the regions near zero and one (Kass 1989), which is where most of our posterior estimates for β_j reside. The function that maps the weights to angles that subtend arcs in the curved space that properly gauge the "information distances" between points on the Jeffreys prior is $\frac{2}{\pi} sin^{-1}\sqrt{\beta}$. Table A-3 therefore reports the transformed mean, 2.5%, and 97.5% estimates of β_j and the accompanying $\hat{R}$.

Table A-3: β_j Diagnostics for Model 1

	mean	2.5%	97.5%	$\hat{R}$
Military Service	0.977	0.948	0.999	1.000
No Combat	0.972	0.936	0.999	1.001
Combat	0.963	0.916	0.999	0.999
War Win	0.969	0.932	0.999	1.000
War Loss	0.946	0.878	0.998	0.999
Military Career	0.943	0.870	0.998	1.000
Military Education	0.885	0.743	0.996	1.000
Rebel	0.807	0.612	0.993	1.001
Rebel Win	0.975	0.943	0.999	0.999
Rebel Loss	0.872	0.731	0.994	1.001
Irregular Entry	0.951	0.889	0.999	1.000

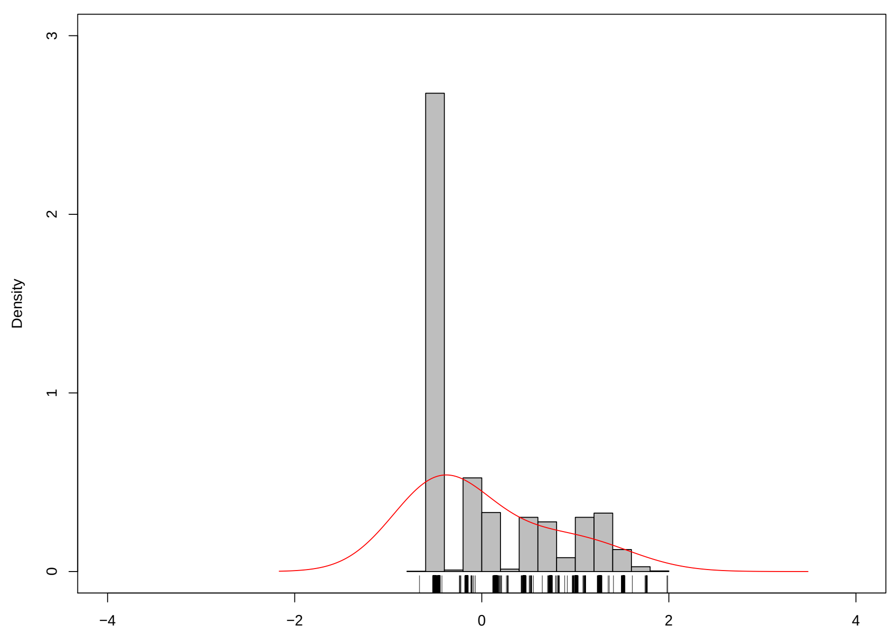

Figure A-1: Distribution of θ_1

Table A-4: Summary Statistics for θ_1

	mean	s.d.	min	max	n
θ_1	0.00	0.66	-0.67	1.99	2965

Figure A-2 report a set of item characteristic curves (ICCs) with 95% credible intervals for Model 1.[1] ICCs plot the probability that a leader possesses a given characteristic (y-axis) as a function of leaders' latent willingness to use military force (x-axis). Item Characteristic Curves incorporate information about both the difficulty (α_j) and discrimination parameters (β_j) in a model, with the steepness of a curve reflecting the relative discriminatory power of a characteristic and the location or height of a curve reflecting the relative difficulty of an item. For example, Figure A-2 tells us that whether a leader served in the military (Column 1, Row 1) does a better job at discriminating among hawkish and dovish leaders than whether a leader was a rebel (Column 2, Row 3). Overall, Figure A-2 provides some face validity to using latent variable techniques to measure leaders' underlying hawkishness; all else equal, possessing various attributes associated with being in the military, participating in a rebellion, or obtaining power through irregular means is associated with a leader's latent willingness to use force.

[1]The code used to estimate and plot the item characteristic curves was adapted from Terechshenko (2017).

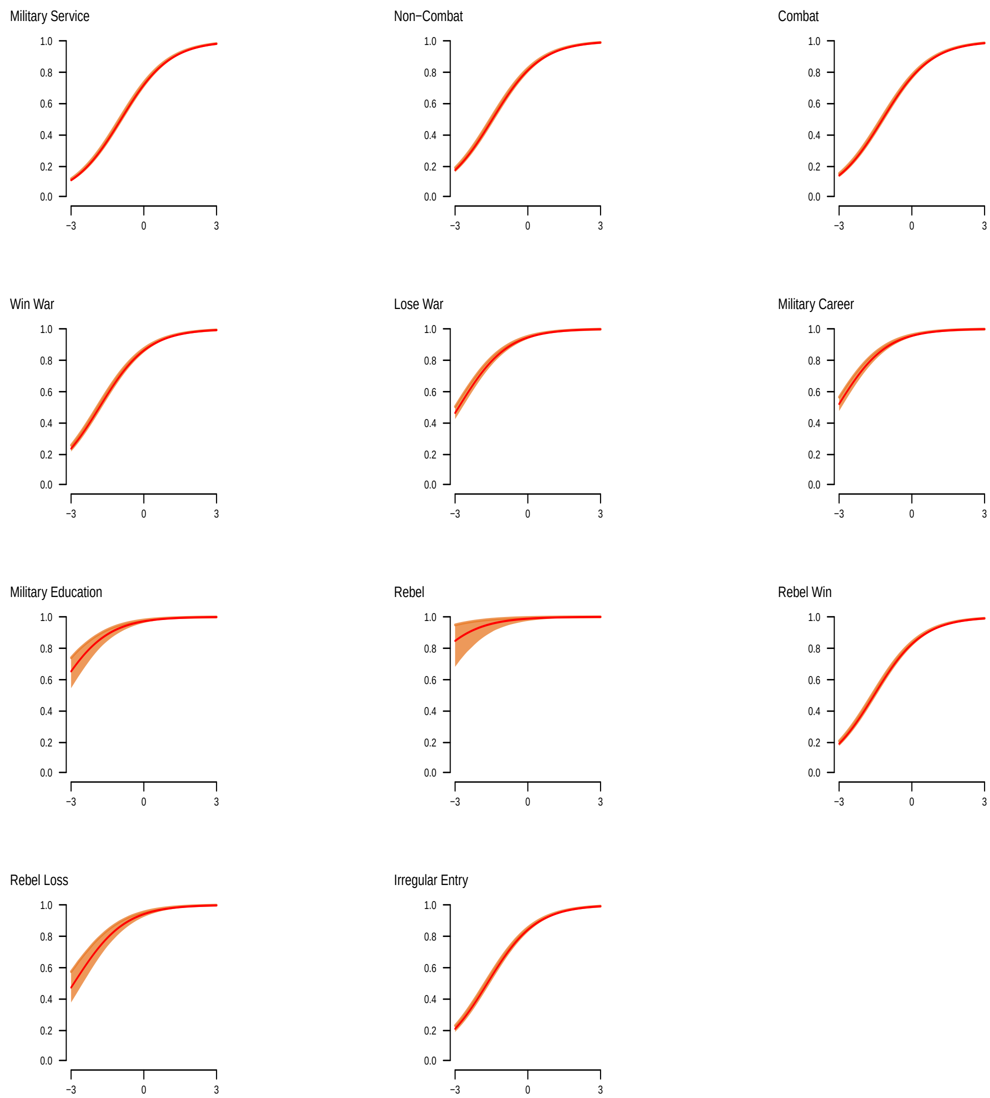

Figure A-2: Item Characteristic Curves for Model 1

C Model 2

Model 2 (M2) expands M1 by including political orientation and psychological characteristic variables/information and assumes $\mathbf{X}$ is a matrix and ϵ a vector:

$$\begin{gathered} Pr(Y_{ij} = 1) = logit^{-1}\beta_j(\theta_i - \alpha_j) \\ \alpha \sim \mathbf{N}\ (0, 10) \\ \beta \sim \mathbf{Beta}\ (\tfrac{1}{2}, \tfrac{1}{2}) \\ \theta \sim \mathbf{N}\ (0, 1) \\ \theta_i \sim (\mathbf{X}\ \epsilon + \upsilon) \end{gathered} \tag{A-1}$$

where $Pr(Y_{ij} = 1)$ is the probability that the ith leader ($n = 2965$) has the jth characteristic ($J = 11$) based on the first eleven indicators in the first column of Table 1 in the main manuscript from the LEAD project (Horowitz, Stam and Ellis 2015) , α represents the difficulty parameter, β represents the discrimination parameter, $\mathbf{X}$ is a 2965 x 6 matrix that contains information on Hitler and the Dalai Lama and from Seki and Williams (2014), Brambor, Lindvall and Stjernquist (2017), and Keller (2005), ϵ is a 6 x 1 vector of coefficients assumed to be unit normal, and υ is a normally distributed error term.

C.1 Diagnostic and Summary Statistics

The following tables report diagnostic and summary statistics associated with α_j and β_j, summary statistics for θ_i and a set of item characteristic curves.

Table A-5: α_j Diagnostics for Model 2

	mean	sd	2.5%	97.5%	$\hat{R}$
Military Service	0.939	0.044	0.853	1.022	1.000
No Combat	1.465	0.050	1.365	1.562	1.000
Combat	1.202	0.047	1.112	1.296	0.999
War Win	1.829	0.054	1.724	1.937	1.000
War Loss	2.843	0.078	2.690	2.999	0.999
Military Career	3.060	0.087	2.897	3.241	1.000
Military Education	3.544	0.169	3.301	3.968	1.001
Rebel	4.575	0.527	4.001	6.002	1.001
Rebel Win	1.559	0.050	1.460	1.655	1.000
Rebel Loss	2.770	0.127	2.581	3.084	1.001
Irregular Entry	1.667	0.054	1.565	1.772	1.000

Table A-6 reports the transformed mean, 2.5%, and 97.5% estimates of β_j and the accompanying $\hat{R}$ for Model 2.

Table A-6: β_j Diagnostics for Model 2

	mean	2.5%	97.5%	$\hat{R}$
Military Service	0.977	0.947	0.999	1.000
No Combat	0.972	0.936	0.999	0.999
Combat	0.962	0.918	0.999	0.999
War Win	0.969	0.931	0.999	1.000
War Loss	0.946	0.880	0.998	1.001
Military Career	0.943	0.871	0.998	1.000
Military Education	0.884	0.747	0.997	1.001
Rebel	0.801	0.606	0.991	1.001
Rebel Win	0.975	0.944	0.999	1.000
Rebel Loss	0.879	0.749	0.997	1.002
Irregular Entry	0.952	0.894	0.998	0.999

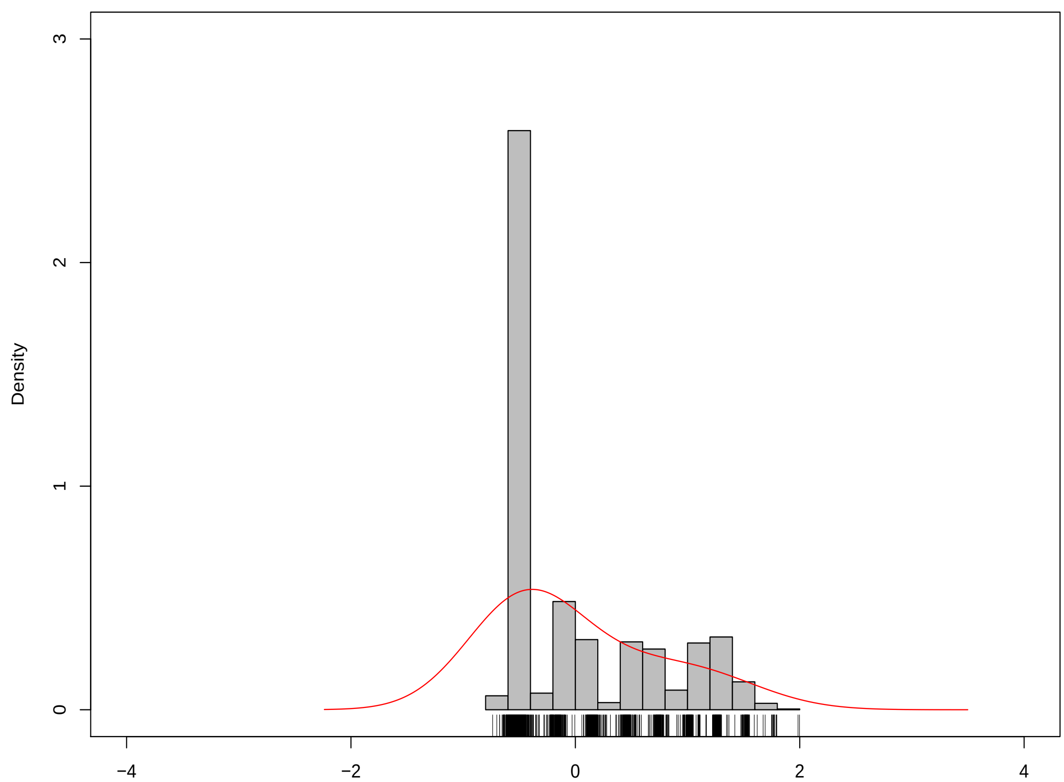

Figure A-3: Distribution of θ_2

Table A-7: Summary Statistics for θ_2

	mean	s.d.	min	max	n
θ_2	0.00	0.66	-0.74	2.00	2965

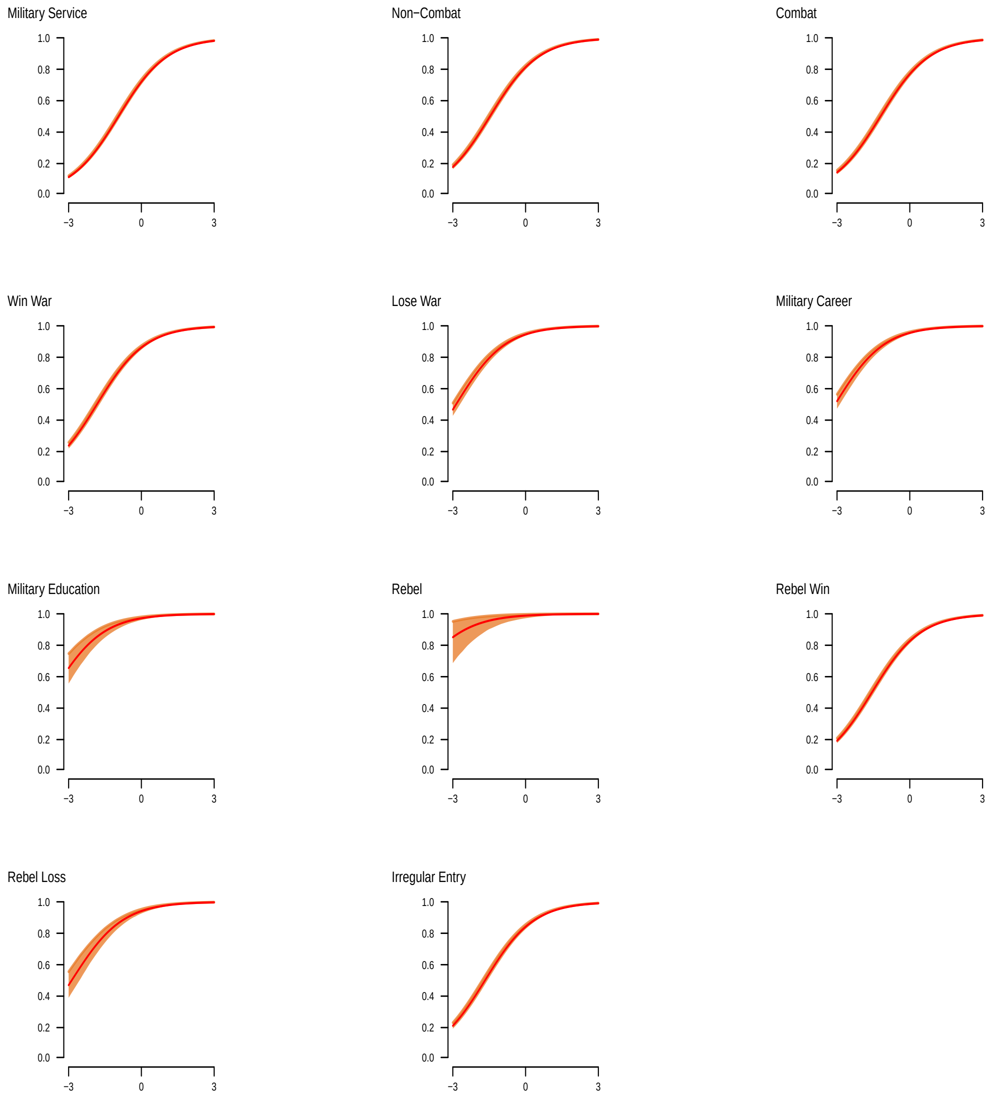

Figure A-4: Item Characteristic Curves for Model 2

D Model 3

Our third model (M3) is identical to M1 except for the background characteristics upon which it is based:

$$\begin{gathered} Pr(Y_{ij} = 1) = logit^{-1} \beta_j(\theta_i - \alpha_j) \\ \alpha \sim \mathbf{N}\ (0, 10) \\ \beta \sim \mathbf{Beta}\ (\tfrac{1}{2}, \tfrac{1}{2}) \\ \theta \sim \mathbf{N}\ (0, 1) \\ \theta_i \sim (\mathbf{X}\ \epsilon + \upsilon) \end{gathered} \tag{A-2}$$

where $Pr(Y_{ij} = 1)$ is the probability that the ith leader ($n = 2965$) has the jth characteristic ($J = 36$) based on all of the indicators from the LEAD project (Horowitz, Stam and Ellis 2015) in Table 1 in the main manuscript, α represents the difficulty parameter, β represents the discrimination parameter, $\mathbf{X}$ is coded +1 for Hitler, -1 for the Dalai Lama, and 0 for all other leaders, ϵ is a coefficient assumed to be unit normal, and υ is a normally distributed error term.

D.1 Diagnostic and Summary Statistics

The following tables report diagnostic and summary statistics associated with α_j and β_j, summary statistics for θ_i and a set of item characteristic curves.

Table A-8: α_j Diagnostics for Model 3

	mean	sd	2.5%	97.5%	$\hat{R}$
Military Service	1.926	0.043	1.841	2.010	1.001
No Combat	3.669	0.136	3.480	4.003	1.001
Combat	2.423	0.048	2.327	2.517	0.999
War Win	3.743	0.077	3.595	3.903	0.999
War Loss	3.955	0.085	3.797	4.126	1.000
Military Career	2.509	0.048	2.419	2.605	0.999
Military Education	2.769	0.053	2.665	2.873	0.999
Rebel	2.173	0.044	2.088	2.262	0.999
Rebel Win	4.396	0.141	4.182	4.743	1.000
Rebel Loss	5.336	0.488	4.873	6.608	1.014
Irregular Entry	2.621	0.052	2.521	2.727	0.999
Male Leader	0.001	0.001	0.000	0.005	1.000
Older Leader	5.451	1.656	2.325	8.841	0.999
Education Level	7.583	1.280	4.999	9.757	0.999
Number of Spouses	0.003	0.003	0.000	0.012	0.999
Married	6.997	1.160	4.973	9.404	1.002
Married in Power	8.079	1.034	6.012	9.855	1.000
Divorced	0.002	0.002	0.000	0.009	0.999
Number of Children	0.071	0.060	0.002	0.227	1.000
Parental Status	6.610	1.148	4.611	9.095	1.000
Illegitimate Child	0.001	0.001	0.000	0.004	1.000
Royalty	0.002	0.002	0.000	0.006	0.999
Orphan	0.001	0.001	0.000	0.005	1.000
Teacher	9.102	0.664	7.557	9.967	1.001
Journalist	8.959	0.706	7.361	9.946	1.000
Lawyer	0.046	0.032	0.002	0.119	1.000
Medical Field	0.001	0.001	0.000	0.005	0.999
Clergy	8.269	0.988	6.267	9.888	1.001
Activist	8.322	0.948	6.378	9.879	1.001
Career Politician	0.044	0.035	0.001	0.129	1.000
Creative	8.943	0.726	7.315	9.942	1.000
Businessman	0.003	0.003	0.000	0.010	0.999
Aristocrat	0.002	0.002	0.000	0.006	1.000
Police	0.001	0.001	0.000	0.004	1.000
Science/Engineer	8.748	0.813	6.952	9.932	0.999
Blue Collar	8.012	1.023	6.029	9.826	1.002

Table A-9 reports the transformed mean, 2.5%, and 97.5% estimates of β_j and the accompanying $\hat{R}$ for Model 3.

Table A-9: β_j Diagnostics for Model 3

	mean	2.5%	97.5%	$\hat{R}$
Military Service	0.974	0.941	0.999	1.000
No Combat	0.884	0.746	0.996	1.001
Combat	0.969	0.930	0.999	1.000
War Win	0.945	0.878	0.998	1.000
War Loss	0.941	0.869	0.997	0.999
Military Career	0.972	0.939	0.999	0.999
Military Education	0.966	0.922	0.999	1.000
Rebel	0.963	0.916	0.998	1.000
Rebel Win	0.903	0.785	0.997	1.000
Rebel Loss	0.833	0.631	0.995	1.010
Irregular Entry	0.951	0.890	0.998	1.000
Male Leader	0.982	0.960	0.999	1.000
Older Leader	0.079	0.009	0.130	1.000
Education Level	0.094	0.062	0.123	1.000
Number of Spouses	0.970	0.931	0.999	1.000
Married	0.501	0.396	0.661	1.002
Married in Power	0.387	0.335	0.468	1.000
Divorced	0.975	0.944	0.999	1.000
Number of Children	0.472	0.409	0.544	0.999
Parental Status	0.554	0.424	0.794	1.001
Illegitimate Child	0.981	0.957	1.000	1.000
Royalty	0.980	0.956	0.999	1.000
Orphan	0.982	0.959	0.999	1.000
Teacher	0.330	0.307	0.371	1.001
Journalist	0.402	0.370	0.459	1.000
Lawyer	0.916	0.823	0.997	1.001
Medical Field	0.981	0.957	0.999	1.000
Clergy	0.546	0.466	0.689	1.001
Activist	0.361	0.320	0.428	1.000
Career Politician	0.802	0.711	0.970	1.000
Creative	0.411	0.377	0.472	1.000
Businessman	0.973	0.940	0.999	1.000
Aristocrat	0.980	0.956	0.999	1.000
Police	0.982	0.959	0.999	1.000
Science/Engineer	0.443	0.400	0.520	0.999
Blue Collar	0.405	0.348	0.493	1.001

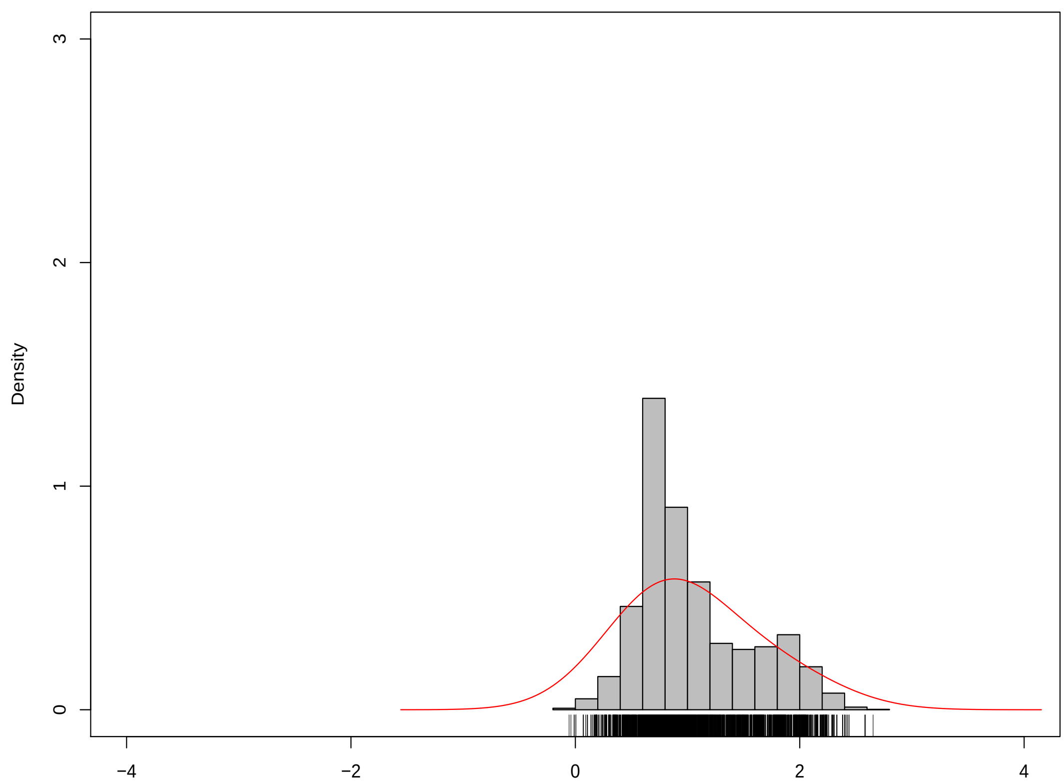

Figure A-5: Distribution of θ_3

Table A-10: Summary Statistics for θ_3

	mean	s.d.	min	max	n
θ_3	1.04	0.49	-0.06	2.65	2965

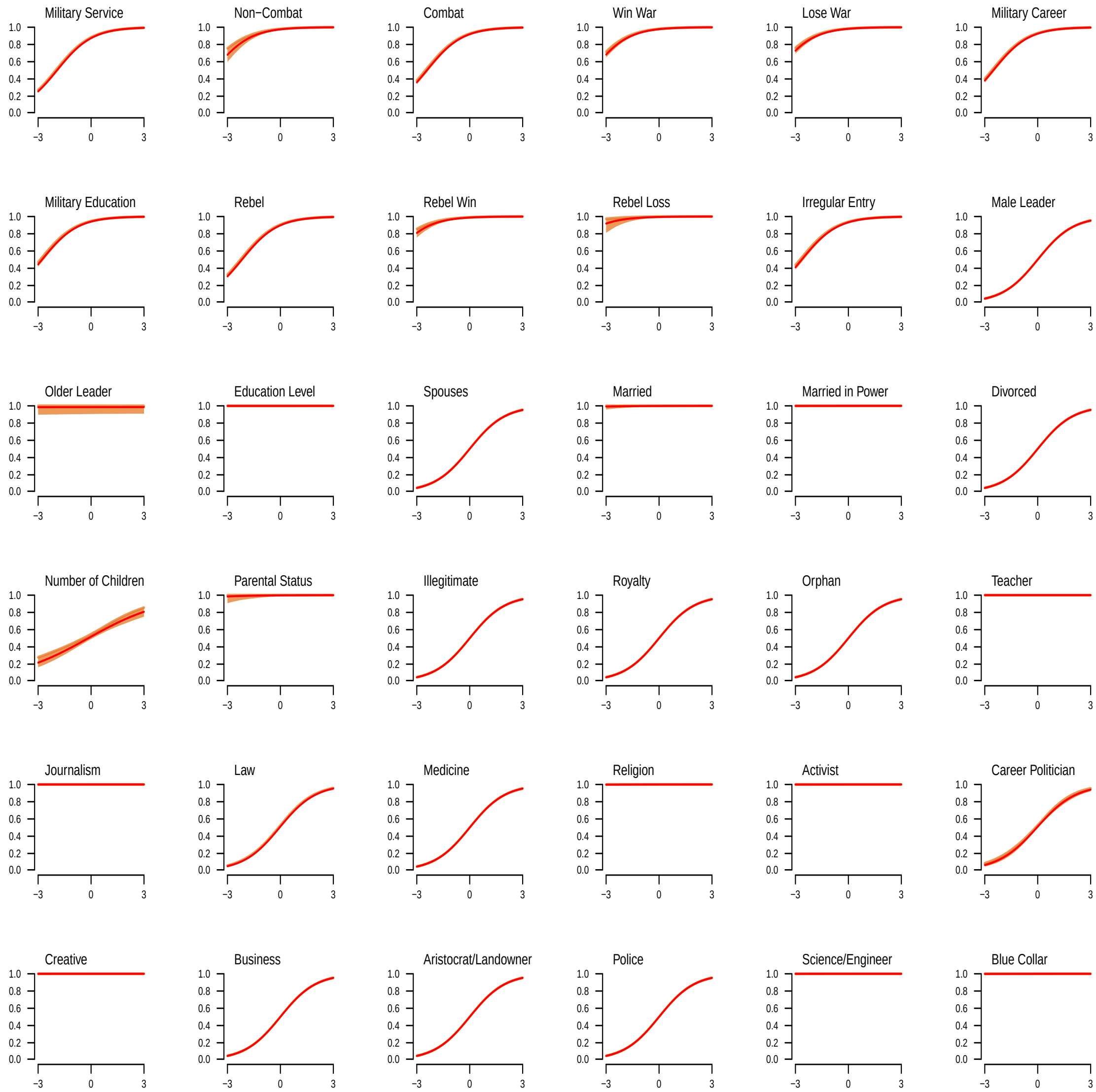

Figure A-6: Item Characteristic Curves for Model 3

E Model 4

Our fourth model (M4) is identical to M2 except for the background characteristics upon which it is based:

$$\begin{gathered} Pr(Y_{ij} = 1) = logit^{-1}\beta_j(\theta_i - \alpha_j) \\ \alpha \sim \mathbf{N}\ (0, 10) \\ \beta \sim \mathbf{Beta}\ (\tfrac{1}{2}, \tfrac{1}{2}) \\ \theta \sim \mathbf{N}\ (0, 1) \\ \theta_i \sim (\mathbf{X}\ \epsilon + \upsilon) \end{gathered} \tag{A-3}$$

where $Pr(Y_{ij} = 1)$ is the probability that the ith leader ($n = 2965$) has the jth characteristic ($J = 11$) based on all of the indicators from the LEAD project (Horowitz, Stam and Ellis 2015) in Table 1 in the main manuscript, α represents the difficulty parameter, β represents the discrimination parameter, $\mathbf{X}$ is a 2965 x 6 matrix that contains information on Hitler and the Dalai Lama and from Seki and Williams (2014), Brambor, Lindvall and Stjernquist (2017), and Keller (2005), ϵ is a 6 x 1 vector of coefficients assumed to be unit normal, and υ is a normally distributed error term.

E.1 Diagnostic and Summary Statistics

The following tables report diagnostic and summary statistics associated with α_j and β_j, summary statistics for θ_i and a set of item characteristic curves.

Table A-11: α_j Diagnostics for Model 4

	mean	sd	2.5%	97.5%	$\hat{R}$
Military Service	1.926	0.042	1.843	2.009	0.999
No Combat	3.666	0.127	3.491	3.987	1.000
Combat	2.423	0.048	2.329	2.517	1.000
War Win	3.743	0.078	3.597	3.898	1.000
War Loss	3.955	0.087	3.792	4.138	0.999
Military Career	2.510	0.048	2.415	2.605	0.999
Military Education	2.769	0.055	2.667	2.876	0.999
Rebel	2.174	0.045	2.087	2.261	1.000
Rebel Win	4.402	0.147	4.176	4.773	1.000
Rebel Loss	5.326	0.416	4.873	6.500	1.002
Irregular Entry	2.622	0.051	2.524	2.725	0.999
Male Leader	0.001	0.001	0.000	0.005	0.999
Older Leader	5.400	1.733	2.162	8.927	0.999
Education Level	7.599	1.288	4.960	9.770	1.000
Number of Spouses	0.004	0.003	0.000	0.012	0.999
Married	6.979	1.112	5.009	9.340	1.000
Married in Power	8.094	1.016	6.028	9.847	1.000
Divorced	0.002	0.002	0.000	0.009	1.000
Number of Children	0.071	0.063	0.002	0.229	1.000
Parental Status	6.642	1.133	4.759	9.133	1.000
Illegitimate Child	0.001	0.001	0.000	0.005	0.999
Royalty	0.002	0.002	0.000	0.006	0.999
Orphan	0.001	0.001	0.000	0.005	1.000
Teacher	9.116	0.638	7.629	9.960	1.001
Journalist	8.968	0.721	7.332	9.948	1.001
Lawyer	0.046	0.031	0.002	0.117	1.000
Medical Field	0.001	0.001	0.000	0.005	0.999
Clergy	8.223	1.019	6.213	9.864	1.000
Activist	8.340	0.949	6.452	9.876	1.000
Career Politician	0.043	0.035	0.002	0.131	0.999
Creative	8.935	0.742	7.264	9.962	1.000
Businessman	0.003	0.003	0.000	0.010	1.000
Aristocrat	0.002	0.002	0.000	0.006	1.000
Police	0.001	0.001	0.000	0.004	0.999
Science/Engineer	8.746	0.800	7.000	9.935	0.999
Blue Collar	7.987	1.077	5.896	9.875	1.001

Table A-12 reports the transformed mean, 2.5%, and 97.5% estimates of β_j and the accompanying $\hat{R}$ for Model 4.

Table A-12: β_j Diagnostics for Model 4

	mean	2.5%	97.5%	$\hat{R}$
Military Service	0.974	0.941	0.999	1.000
No Combat	0.886	0.752	0.996	1.000
Combat	0.969	0.930	0.999	1.000
War Win	0.946	0.876	0.998	1.000
War Loss	0.941	0.865	0.998	1.000
Military Career	0.973	0.940	0.999	1.000
Military Education	0.966	0.924	0.999	1.003
Rebel	0.962	0.917	0.999	1.000
Rebel Win	0.901	0.779	0.997	1.000
Rebel Loss	0.832	0.642	0.994	1.003
Irregular Entry	0.951	0.892	0.998	1.000
Male Leader	0.982	0.959	0.999	1.000
Older Leader	0.078	0.009	0.129	1.001
Education Level	0.094	0.063	0.121	1.000
Number of Spouses	0.969	0.931	0.999	1.000
Married	0.501	0.398	0.656	1.000
Married in Power	0.386	0.335	0.469	1.000
Divorced	0.975	0.945	0.999	1.000
Number of Children	0.472	0.405	0.544	1.000
Parental Status	0.551	0.422	0.747	1.000
Illegitimate Child	0.981	0.958	0.999	1.000
Royalty	0.979	0.953	0.999	1.000
Orphan	0.981	0.957	0.999	1.000
Teacher	0.329	0.307	0.369	1.001
Journalist	0.402	0.370	0.461	1.001
Lawyer	0.915	0.819	0.997	1.000
Medical Field	0.981	0.956	0.999	0.999
Clergy	0.550	0.468	0.698	1.000
Activist	0.361	0.319	0.427	1.000
Career Politician	0.802	0.710	0.980	1.000
Creative	0.411	0.376	0.474	1.000
Businessman	0.973	0.940	0.999	0.999
Aristocrat	0.980	0.956	0.999	1.000
Police	0.982	0.960	0.999	0.999
Science/Engineer	0.443	0.400	0.517	0.999
Blue Collar	0.407	0.348	0.499	1.000

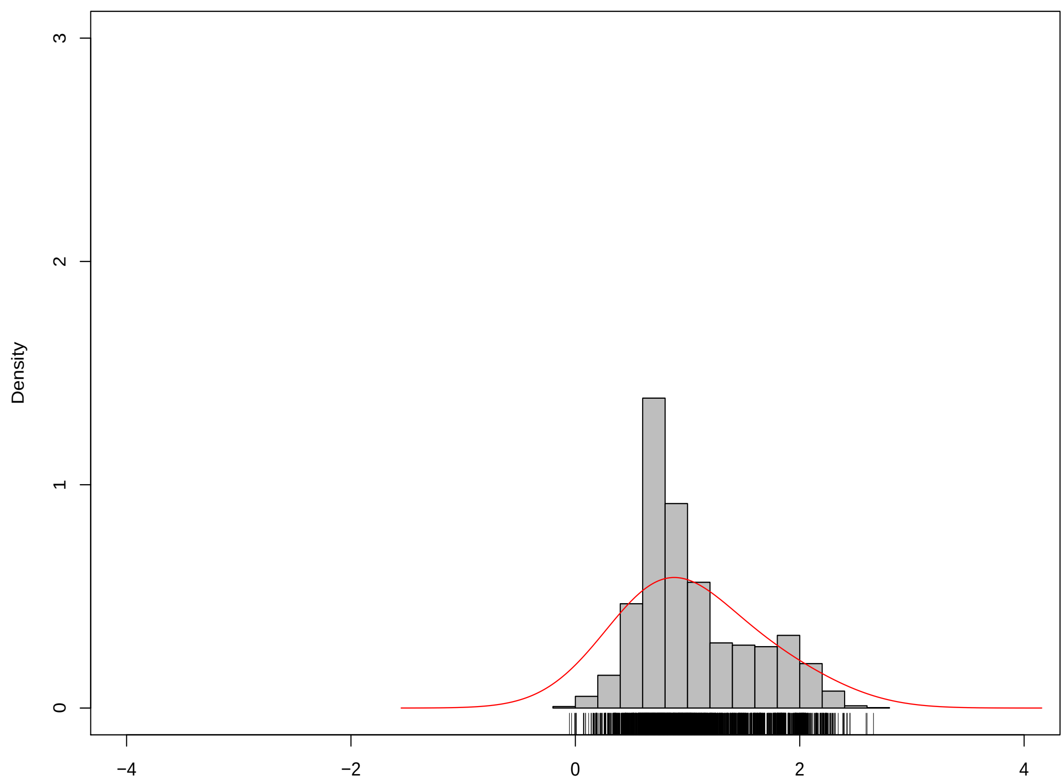

Figure A-7: Distribution of θ_4

Table A-13: Summary Statistics for θ_4

	mean	s.d.	min	max	n
θ_4	1.04	0.49	-0.05	2.66	2965

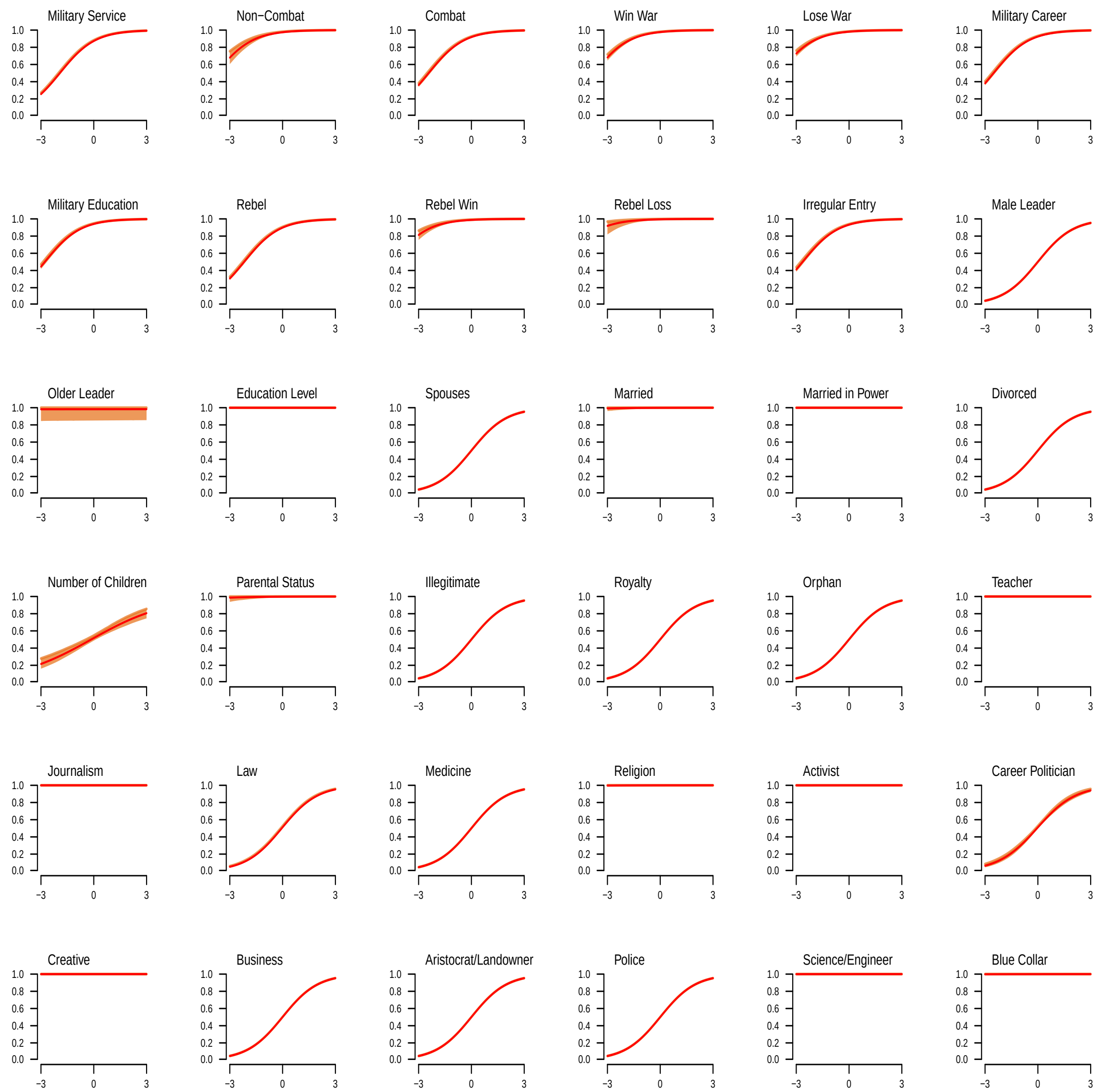

Figure A-8: Item Characteristic Curves for Model 4

2 Validation Results

This section presents the results of the models estimated in order to conduct the validation analyses reported in the main manuscript. Tables A-1 - A-3 present results for the initiation of an ICB crisis, the initiation of any MID, and the initiation of a MID in which both sides used force, respectively.

Table A-1: Validation Analyses using ICB Crisis Initiation

	Null	Military Service	M1	M2	M3	M4
(Intercept)	-2.98 *	-3.27 *	-3.14 *	-3.14 *	-3.81 *	-3.83 *
	(0.04)	(0.06)	(0.05)	(0.05)	(0.11)	(0.11)
Prior Military Service		0.67 *				
		(0.09)				
Model 1			0.61 *			
			(0.06)			
Model 2				0.63 *		
				(0.06)		
Model 3					0.71 *	
					(0.08)	
Model 4						0.72 *
						(0.08)
N	10838	10838	10838	10838	10838	10838
AIC	4204.93	4151.06	4102.67	4094.92	4126.28	4123.10
BIC	4234.10	4209.39	4161.00	4153.25	4184.60	4181.42
$\log L$	-2098.47	-2067.53	-2043.33	-2039.46	-2055.14	-2053.55

Standard errors in parentheses.

* indicates significant at $p < 0.05$ with two-tailed test.

Table A-2: Validation Analyses using MID Initiation

	Null	Military Service	M1	M2	M3	M4
(Intercept)	-1.46 *	-1.69 *	-1.54 *	-1.54 *	-2.10 *	-2.11 *
	(0.02)	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)
Prior Military Service		0.55 *				
		(0.05)				
Model 1			0.41 *			
			(0.03)			
Model 2				0.42 *		
				(0.03)		
Model 3					0.57 *	
					(0.04)	
Model 4						0.57 *
						(0.04)
N	11544	11544	11544	11544	11544	11544
AIC	11177.40	11048.29	11020.15	11011.41	11007.66	11003.53
BIC	11206.82	11107.12	11078.98	11070.24	11066.49	11062.37
$\log L$	-5584.70	-5516.14	-5502.07	-5497.70	-5495.83	-5493.77

Standard errors in parentheses.

* indicates significant at $p < 0.05$ with two-tailed test.

Table A-3: Validation Analyses using Severe MID Initiation

	Null	Military Service	M1	M2	M3	M4
(Intercept)	-2.82 *	-3.09 *	-2.96 *	-2.96 *	-3.58 *	-3.59 *
	(0.04)	(0.06)	(0.05)	(0.05)	(0.10)	(0.10)
Prior Military Service		0.62 *				
		(0.08)				
Model 1			0.54 *			
			(0.05)			
Model 2				0.55 *		
				(0.05)		
Model 3					0.65 *	
					(0.07)	
Model 4						0.65 *
						(0.07)
N	11544	11544	11544	11544	11544	11544
AIC	5010.36	4953.58	4909.83	4906.69	4930.26	4929.14
BIC	5039.78	5012.41	4968.66	4965.52	4989.09	4987.97
$\log L$	-2501.18	-2468.79	-2446.92	-2445.35	-2457.13	-2456.57

Standard errors in parentheses.

* indicates significant at $p < 0.05$ with two-tailed test.

3 An Alternative Set of Models

One of the key features of our approach is that it is relatively easy to estimate alternative measures of leaders' latent hawkishness by altering any of the assumptions underlying our models. As mentioned in the main manuscript, one assumption to alter is related to how the various items/background characteristics β_j are weighted in terms of their relative influence on the latent variable θ. Our models assume a Jeffreys prior that binds estimates of $\beta_j \in [0, 1]$ (i.e., $\beta \sim$ **Beta** $(\frac{1}{2}, \frac{1}{2})$). It might be more natural to assume that β_j follows a lognormal distribution, which would constrain each weight to be positive but would not impose an upper bound on the estimated weight or influence of any given background characteristic. We therefore estimated a set of models that were otherwise identical to M1-M4 discussed in the main manuscript, except for the assumption that $\beta \sim lognormal(0, 2)$.

We think there are three important things to note about the resulting models. First, whether a leader previously served in the military has a substantially larger effect on the latent variable than any other item/background characteristic. To demonstrate this, Table A-1 reports our estimates of β_j yielded from a model that corresponds to M1 in the main manuscript ($M1_{ln}$). The estimated discrimination parameter associated with *Military Service* is 192, or roughly nine times larger than the next to largest discrimination parameter. Indeed, the discrimination parameter associated with *Military Service* is at least seven times larger than the next largest discrimination parameter in each of the four alternative models.

Table A-1: β_j Diagnostics for $M1_{ln}$

	mean	2.5%	97.5%	$\hat{R}$
Military Service	191.58	49.73	720.65	1.01
No Combat	6.53	5.79	7.43	1.00
Combat	2.08	1.86	2.31	1.00
War Win	4.75	4.21	5.32	1.00
War Loss	3.10	2.66	3.57	1.00
Military Career	2.80	2.37	3.25	1.00
Military Education	1.24	0.94	1.58	1.00
Rebel	1.05	0.67	1.47	1.00
Rebel Win	20.57	13.05	38.60	1.01
Rebel Loss	1.90	1.62	2.19	1.00
Irregular Entry	1.87	1.64	2.09	1.00

Second, while we encountered no issues with the models reported in the main manuscript, diagnostic analyses identified estimation issues with each of the models. First, each model had multiple divergent transitions after the 1000 warmup iterations. Second, models $M1_{ln}$ and $M2_{ln}$ had a large number of iterations that exceeded the standard maximum tree depth specified by STAN (998 and 1641, respectively). Third, each of the models had multiple parameters with an effective sample size less than 10% of the total sample size. All of these issues are indicative of poor mixing between the likelihood (data) and the lognormal prior. If we found that the thetas from lognormal priors on beta stood out favorably in terms of predictive validity, then we might continue tuning the model and increase iterations to see if we could make the mixing problems go away. However, it is clear that the reverse is true. Moreover, given that we ran over 4000 post-warmup iterations with a high target proposal acceptance probability (the default in rstan is .8; we used .95), we regard the lognormal results with suspicion.

Third, and we think most importantly, the measures of leaders' latent hawkishness produced by the alternative models are worse predictors of interstate conflict initiation than the measures described in the main manuscript. This statement is based on two sets of comparisons. The first was analogous to the validation analyses reported in the main manuscript and analyzed how much better the alternative measures predicted conflict initiation than a measure of whether a leader previously served in the military. The second set of analyses compared the ability of each of the alternative measures to predict conflict initiation to the analogous measure described in the main manuscript (M1 vs. $M1_{ln}$, M2 vs. $M2_{ln}$, etc.).

Table A-2 reports a set of Vuong statistics associated with pairwise comparisons of bivariate logits using the respective measures that allow us to make the two sets of relevant comparisons for each of our three measures of conflict initiation. Positive and significant Vuong statistics indicate the "row" model performed better than the "column" model while negative and significant Vuong statistics indicate the "column" model performed better than the "row" model. Focusing on the first row of each panel, we find that, compared to prior military service, $M1_{ln}$ and $M2_{ln}$ are significantly stronger predictors of conflict initiation but $M3_{ln}$ and $M4_{ln}$ are not. In comparison, M1-M4 all explain more variation in conflict initiation than prior military service (see validation analyses reported in the main manuscript). The remaining cell entries in each panel report the results of pairwise comparisons between our main measures (rows) and alternative measures (columns). The Vuong statistic in each of these cells is positive and it is statistically significant in eleven of the twelve cells. Thus, with the exception of the M1 vs. $M1_{ln}$ comparison when predicting the initiation of any MID, the measures reported in the main manuscript perform significantly better when predicting our three indicators of interstate conflict initiation than the alternative measures yielded from models that assume $\beta \sim lognormal(0, 2)$.

Table A-2: Vuong Statistics for Assessing Measures' Performance

Panel A: ICB Initiation

	$M1_{ln}$	$M2_{ln}$	$M3_{ln}$	$M4_{ln}$
Military	-4.56**	-4.57**	0.42	0.42
M1	2.25*		–	–
M2	–	3.63**	–	–
M3	–	–	4.03**	–
M4	–	–	–	4.36**

Panel B: MID Initiation

	$M1_{ln}$	$M2_{ln}$	$M3_{ln}$	$M4_{ln}$
Military	-2.51**	-2.55**	-0.73	-0.74
M1	0.50	–	–	–
M2	–	1.91*	–	–
M3	–	–	3.20**	–
M4	–	–	–	3.57**

Panel C: Severe MID Initiation

	$M1_{ln}$	$M2_{ln}$	$M3_{ln}$	$M4_{ln}$
Military	-4.16**	-4.15**	-1.02	-1.05
M1	1.98*	–	–	–
M2	–	2.59**	–	–
M3	–	–	1.99**	–
M4	–	–	–	2.11**

* indicates significance at $p < 0.05$; ** indicates $p < 0.01$.

Why might it be the case that more restrictive models that bind the discrimination parameters between zero and one produce more predictive measures than models that do not place an upper bound on the discrimination parameters? Shortly after we began our research, we recognized that a fundamental difference between the leader background data from Horowitz, Stam and Ellis (2015) and the data IRT models were initially developed to reduce – educational testing data – requires that the discrimination items be confined in our context to a bounded space. The key feature in the data is the prevalence of rows with zeroes on all or most of the items. This feature is easiest to recognize in M1 and M2, where most of the items relate to prior military experience. The baseline item in this model – the one that must be scored "1" in order for many of the others to potentially take on a value of "1" – is the prior military service dummy. Absent an upper bound on β, an IRT model will identify military service as almost-deterministically crucial, and assign it a weight many times that of the others. As noted above, this is what occurred in each of the four alternative models we estimated. This has the effect in the whole data set of making the variance of the latent variable less empirically distinct from that of the dummy for military service, which limits a latent variable's ability to be a better predictor of conflict initiation than the prior military service dummy. Future research might improve upon our measures by experimenting with alternative priors over β, but not placing an upper bound on β appears to limit the predictive strength of latent measures of leaders' willingness to use military force.

References

Brambor, Thomas, Johannes Lindvall and Annika Stjernquist. 2017. "The Ideology of Heads of Government, 1870–2012. Version 1.5." • .

Horowitz, Michael C, Allan C Stam and Cali M Ellis. 2015. *Why Leaders Fight*. Cambridge University Press.

Kass, Robert E. 1989. "The geometry of asymptotic inference." *Statistical Science* pp. 188–219.

Keller, Jonathan W. 2005. "Leadership Style, Regime Type, and Foreign Policy Crisis Behavior: A Contingent Monadic Peace?" *International Studies Quarterly* 49(2):205–232.

Seki, Katsunori and Laron K Williams. 2014. "Updating the Party Government data set." *Electoral Studies* 34:270–279.

Terechshenko, Zhanna. 2017. "Item-Response Theory Model with RStan." http://zterechshenko.com/2017/08/09/a-post-about-my-research.html.