Online Appendix for "The Prospects of Punishment and the Strategic Escalation of Civil Conflicts" (Not for publication)

February 14, 2023

Contents

A Estimation details

Recall that the equilibrium of the model can be described as a system of three equations $\Psi(p^*;\theta)$, and p^* is an equilibrium only if $p^* = \Psi(p^*;\theta)$. We now write out the full form of this system of equations:

$$p_c^* = 1 - \Phi\left(\frac{S_R - (1-p_g^*)V_R}{p_g^*} - \bar{W}_R\right)\Phi\left(\frac{S_R - (1-p_g^*)V_R}{p_g^*} - \bar{r}\right) \equiv f(p_g^*;\theta), \tag{A.1}$$

$$p_e^* = \Phi_2\left(\frac{\bar{W}_R - \bar{r}}{\sqrt{2}}, \bar{W}_R - \frac{S_R - (1-p_g^*)V_R}{p_g^*}; \frac{1}{\sqrt{2}}\right) \Big/ f(p_g^*;\theta) \equiv h(p_g^*;\theta), \tag{A.2}$$

and

$$p_g^* = \Phi\left(\frac{h(p_g^*;\theta)\bar{W}_G + (1-h(p_g^*;\theta))V_G - C_G}{h(p_g^*;\theta)}\right) \equiv g\left(h(p_g^*;\theta)\right). \tag{A.3}$$

In the above, $\Phi(x)$ is the standard normal cumulative distribution function (CDF) and $\Phi_2(x,y;\rho)$ is the standard bivariate normal CDF ($\sigma_x^2 = \sigma_y^2 = 1$) with correlation ρ. Notice that p_g^* completely pins down the equilibrium: Equations A.1 and A.2 are R's best responses to p_g^*, while Equation A.3 is G's best response to $h(p_g^*;\theta)$. Because this is not an original model, I do not derive these expressions of the choice probabilities except to note the probabilities are all formed based on ordinary comparison of the expected utility of an action compared to private information as in nearly all random utility models. More thorough discussions of the derivations can be found in Lewis and Schultz (2003) or Jo (2011).

Let $y_d \in \{SQ, CL, BD, CC\}$ denote the outcome of a group and government dyadic observation $d = 1, \ldots, D$. To estimate the parameter vector β, we start by constructing the multinomial log-likelihood

$$\begin{aligned} L(\beta|y) = \sum_{d=1}^{D} &\mathbb{I}(y_d = SQ)\log\left[1 - f(p_{g,d};\beta)\right] + \mathbb{I}(y_d = CL)\log\left[f(p_{g,d};\beta)(1 - g(h(p_{g,d};\beta)))\right] \\ &+ \mathbb{I}(y_d = BD)\log\left[f(p_{g,d};\beta)g(h(p_{g,d};\beta))(1 - h(p_{g,d};\beta))\right] \\ &+ \mathbb{I}(y_d = CC)\log\left[f(p_{g,d};\beta)g(h(p_{g,d};\beta))h(p_{g,d};\beta)\right], \end{aligned} \tag{A.4}$$

where $p_{g,d}$ is found by solving Eq. A.3 for each observation at every guess of β and $\mathbb{I}(\cdot)$ is the indicator function. Ideally, we want to find the values of β that maximize this log-likelihood. However, it is well known that this game admits multiple equilibria under many reasonable sets of parameters (Crisman-Cox and Gibilisco 2021; Jo 2011), which makes this task less than straightforward.

In particular, Crisman-Cox and Gibilisco (2021) demonstrate that Eq. A.4 is not well defined because multiple equilibria are possible and using Eq. A.4 is problematic for estimation *even if there is a unique equilbrium at the true parameters values*. Specifically, this issue means that any given guess of the parameter vector could be consistent with multiple values of p_r for each observation and can consequently produce multiple log-likelihood values (which is to say that the above log-likelihood is a correspondence rather than a function). As a result of this multiplicity, standard optimization techniques will frequently converge to incorrect estimates (incorrect in the sense of not actually maximizing Eq. A.4). They also note that standard refinements, such as the Intuitive Criterion or regularity, do not provide any relief here as all equilibria of this game survive these refinements. Likewise, while other ad-hoc refinements can be employed to eliminate the multiplicity, they will introduce severe discontinuities into the above log-likelihood and make direct optimization of Eq. A.4 substantially less feasible. The number of discontinuities tends to increase with the number of observations, compounding these problems in datasets of any meaningful size.

As a result of these concerns I use the nested pseudo-likelihood (NPL) estimator they propose to find the parameters of interest.[1] This estimation procedure sidesteps the problems associated with multiple equilibria by assuming that equilibrium selection is a function of the observed covariates. To put this another way, this approach imposes an equilibrium selection rule that is empirical, rather than theoretical, and allows the data to tell us which equilibrium the actors reach. The estimation routine proceeds as follows:

1. The equilibrium choice probabilities $p_{g,d}$ and $p_{e,d}$ are estimated using a flexible method (in this case a random forest) that relates the decision nodes to all the observed covariates. These initial estimates of equilibrium behavior need not, and likely will not, satisfy the equilibrium conditions in Equations 1-3, but this is not an issue as these probabilities only serve as an initial guess in an iterative process.

2. Estimate β by maximizing the log-pseudo-likelihood with the equilibrium quantities fixed to

[1]Specifically, I use their `R` package (`sigInt`) for estimation and analysis.

the current estimates of $p_{g,d}$ and $p_{e,d}$.

3. Using the estimates of β from step 2, update the estimates of $p_{g,d}$ and $p_{e,d}$ using best-response functions g and h, respectively.

4. Iterate steps 2 and 3 until convergence.

The log-pseudo-likelihood from step 2 is given by:

$$\begin{aligned} PL(\beta|y,\hat{p}) = \sum_{d=1}^{D} & \mathbb{I}(y_d = SQ)\log\left[1 - f(\hat{p}_{g,d};\beta)\right] + \mathbb{I}(y_d = CL)\log\left[f(\hat{p}_{g,d};\beta)(1 - g(\hat{p}_{e,d};\beta))\right] \\ & + \mathbb{I}(y_d = BD)\log\left[f(\hat{p}_{g,d};\beta)g(\hat{p}_{e,d};\beta)(1 - h(\hat{p}_{g,d};\beta))\right] \\ & + \mathbb{I}(y_d = CC)\log\left[f(\hat{p}_{g,d};\beta)g(\hat{p}_{e,d};\beta)h(\hat{p}_{g,d};\beta)\right], \end{aligned} \tag{A.5}$$

where $\hat{p} = (\hat{p}_{e,d}, \hat{p}_{g,d})_{d=1}^{D}$ refers to the current estimates of the choice probabilities from steps 1 and 3. Notice that the equilibrium quantity $p_{g,d}$ is no longer endogenously defined and so it does not have to computed at every optimization step; this means that the indeterminacies/discontinuities in Eq. A.4 are not present in the pseudo-likelihood function. The intuition behind this approach is that if we know the true equilibrium choice probabilities we could fix $p_{g,d}$ to these values in Eq. A.4, which would turn it into a well-behaved and continuous function. However, since we do not know these choice probabilities we estimate them from the observables to generate a feasible estimator.[2] The iterative process means that the NPL estimates will be in equilibrium at convergence and overall the estimation routine will be both better and faster than trying to directly optimize Eq. A.4. For more details see Crisman-Cox and Gibilisco (2021).

Except where listed, the models in the main text and Online Appendix use the full NPL algorithm. Models that are labeled pseudo-likelihood (PL) below only do steps 1-2 from the algorithm. Crisman-Cox and Gibilisco (2021) demonstrate the PL performs well with this model. In cases where the NPL fails to converge (i.e., the country- and group-fixed-effects models), I use the PL estimator.

[2] The term "feasible" is used here in the same sense as feasible generalized least squares (GLS), where unknown values are estimated in a first step approach to provide traction on a GLS problem.

B Summary statistics

In this appendix, I report summary statistics and sources for all the independent variables used in Models 2 & 4. The former are presented in Table B.1 and latter in Table B.2. The summary statistics reflect the pre-standardized measurements of the variables.

Table B.1: Independent Variable Summaries (Model 2, CONIAS & MAR)

Variable	Min	Mean	Max	Source
Separatist	0.00	0.50	1.00	CONIAS & MARS
GDP pc (logged)	−1.61	0.99	4.68	Penn World Table
Polity2	−10.00	−1.18	10.00	Polity IV
Other groups	0.00	5.04	37.00	CONIAS & MARS
Rough terrain (logged)	2.25	4.81	6.02	Shaver, Carter, & Shawa (2019)
Mil. Per. pc (logged)	0.00	1.75	5.04	COW
Population (logged)	5.38	9.79	14.06	COW
Oil exporter	0.00	0.14	1.00	World Bank

Table B.2: Independent Variable Summaries (Model 4, REVMOD)

Variable	Min	Mean	Max	Source
Separatist	0.00	0.32	1.00	REVMOD
GDP pc (logged)	−1.19	1.37	3.73	Penn World Table
Polity2	−10.00	1.44	10.00	Polity IV
Other groups	0.00	7.19	40.00	REVMOD
Political organization	0.00	0.26	1.00	REVMOD
Rough terrain (logged)	2.34	5.06	6.02	Shaver, Carter, & Shawa (2019)
Mil. Per. pc (logged)	0.11	1.95	4.36	COW
Population (logged)	6.68	10.52	14.00	COW
Oil exporter	0.00	0.17	1.00	World Bank
Group size (logged)	2.30	6.20	14.51	REVMOD
Islamic	0.00	0.26	1.00	REVMOD
Leftist	0.00	0.36	1.00	REVMOD

C Other parameter estimates

In this section, I present the other parameter estimates from the models estimated in Table 1 (main text). These estimates are listed in Table C.1. As mentioned, these estimates reflect the changes to specific payoffs for each actor, but their effect on choice probabilities involves a set of highly nonlinear functions. With that caveat in mind, there are a few points of interest in these estimates.

Table C.1: Structural estimates for the other payoffs of the main models

Outcome	Payoff	Variable	Model 1	Model 2	Model 3	Model 4
Gov't Conciliates (CL)	V_R	Constant	2.21	2.00	1.47	4.62
			(2.94)	(1.32)	(1.02)	(1.98)
		Separatist dispute	0.86	−0.00	−0.30	−2.44
			(1.00)	(0.79)	(0.67)	(2.07)
		GDP pc	−0.64	−0.66	−0.16	−0.82
			(1.10)	(0.45)	(0.35)	(1.56)
	C_G	Constant	−0.25	−1.91	−1.72	1.32
			(3.33)	(2.03)	(1.37)	(0.43)
		Democracy	0.11	0.19	0.14	0.06
			(0.16)	(0.13)	(0.13)	(0.04)
		Separatist dispute	0.11	0.19	0.27	0.30
			(0.40)	(0.24)	(0.29)	(0.19)
		Other groups	0.16	0.09	0.14	−0.95
			(0.16)	(0.12)	(0.08)	(0.17)
		Political wing				1.35
						(0.22)
Civil Conflict (CC)	$\bar{W}_R$	Constant	−1.23	−1.57	−1.53	0.60
			(0.19)	(0.13)	(0.11)	(0.10)
		Rugged terrain	0.09	0.05	0.03	−0.12
			(0.09)	(0.07)	(0.06)	(0.11)
		GDP pc	−0.16	−0.18	−0.19	0.17
			(0.14)	(0.10)	(0.08)	(0.11)
		Military size	−0.15	−0.10	−0.08	−0.00
			(0.29)	(0.10)	(0.07)	(0.09)
		Population	−0.05	−0.01	0.03	0.13
			(0.09)	(0.07)	(0.06)	(0.08)
		Oil Exporter	0.25	0.10	0.10	0.27
			(0.17)	(0.13)	(0.13)	(0.22)
		Group size				−0.49
						(0.10)
	$\bar{W}_G$	Constant	1.19	−0.67	−0.49	4.58
			(3.62)	(2.23)	(1.49)	(0.63)
		Rugged terrain	0.08	0.12	0.09	0.10
			(0.14)	(0.11)	(0.10)	(0.05)
		GDP pc	−0.37	−0.49	−0.52	−0.28
			(0.19)	(0.18)	(0.15)	(0.11)
		Military size	−0.06	−0.14	−0.11	−0.17
			(0.55)	(0.27)	(0.22)	(0.11)
		Population	0.12	0.13	0.29	−0.09
			(0.20)	(0.16)	(0.16)	(0.09)
		Group size				−0.16
						(0.12)
Log Likelihood			–717.13	–897.68	–921.87	–365.66
Observations			1207	2352	2357	533

Standard errors in parentheses. All non-binary covariates are standardized. Constants represent the payoff when non-binary covariates are at their average and dummy variables are 0. Additional model information can be found in Table 1.

In Models 1-3 concessions are on average costly for the government (i.e., less than zero when other variables are held at their means). In Model 4, we find the odd result that the government prefers conceding to victory without fighting in nearly all observations, which suggests that this specification may not fit the data as well as the others. One reason for this abnormality could be the lack of status quo cases and the much smaller sample size. Of further note, concessions are less costly for democracies, on average. Likewise, in Model 4 we find that the presence of a political wing within the group makes concessions more appealing to the government, perhaps because this signals a certain level of seriousness and ability to commit to the deal.

D Episodic data

Table D.1: Punishment estimates $\bar{r}$ for episodic data

	Estimate
Constant	−1.37
	(0.20)
Democracy	0.18
	(0.14)
Military size	0.06
	(0.18)
Other groups	−0.16
	(0.17)
Separatist Dispute	−0.02
	(0.28)
Case selection	CONIAS & MAR
Outcome coding	Researcher
Log-likelihood	−377.38
Observations	394
% $V_R > S_R$	100%
% $V_G > C_G$	100%
% $S_R > \bar{r}$	100%

Standard errors in parentheses

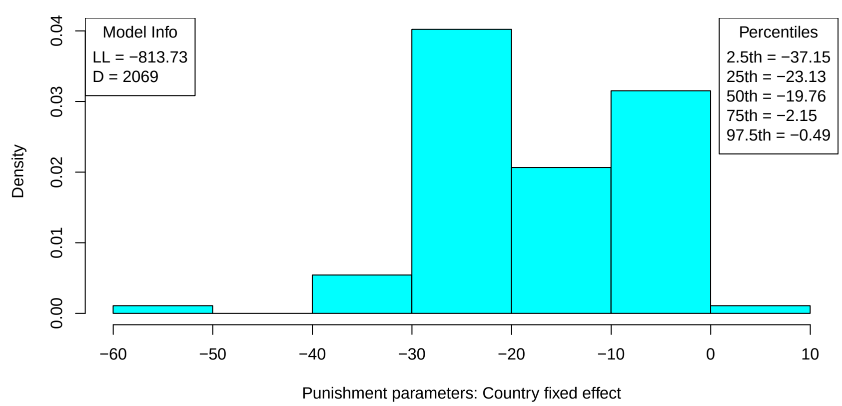
In this appendix, I consider a different approach to aggregating the data. Specifically, here I look at the episodes in terms of a government-group-episode rather than the government-group-decade. Each government-group pair now only enters the data once and only the most extreme outcome of the interaction is considered. The main effect here is to reduce or limit the number of status quo observations. In the interest of space, I only present the coefficients associated with $\bar{r}$ as these are the main parameters of interest. The rest of the model remains specified with the same controls as Model 2.

The results from the episodic data are reported in Table D.1. Here we see that the estimates are largely unchanged, although there are some minor differences. Most importantly however, the punishment parameters in $\bar{r}$ are roughly the same magnitude and direction. As with the dyad-decade data, we see that on average groups expect to be punished for backing down in a dispute.

E Fixed-effects approaches

In this appendix, I consider country-specific and then group-specific fixed-effects approaches to specifying the punishment parameters $\bar{r}$. The pseudo-likelihood routines exhibited signs of numerical instability as the number of parameters gets increasingly large, making standard error estimation unreliable. As such, no standard errors are reported; the values in this section are best thought of as a model calibration exercise.

Figure E.1: Calibrated estimates of $\bar{r}$ with country fixed effects (CONIAS)



The country-specific punishments are presented in Figure E.1, while the group-specific results are presented in Figures E.2 (CONIAS) and E.3 (REVMOD). A few interesting results are apparent from looking across these three figures. First, the estimates are overwhelmingly negative, further supporting the findings in the main text. For almost all groups backing down is, on average, a costly proposition. Indeed, for the REVMOD cases, we see that the anticipated punishments are uniformly negative, while in the other cases there are only a few exceptions. Second, there is

Figure E.2: Calibrated estimates of $\bar{r}$ with group fixed effects (CONIAS)

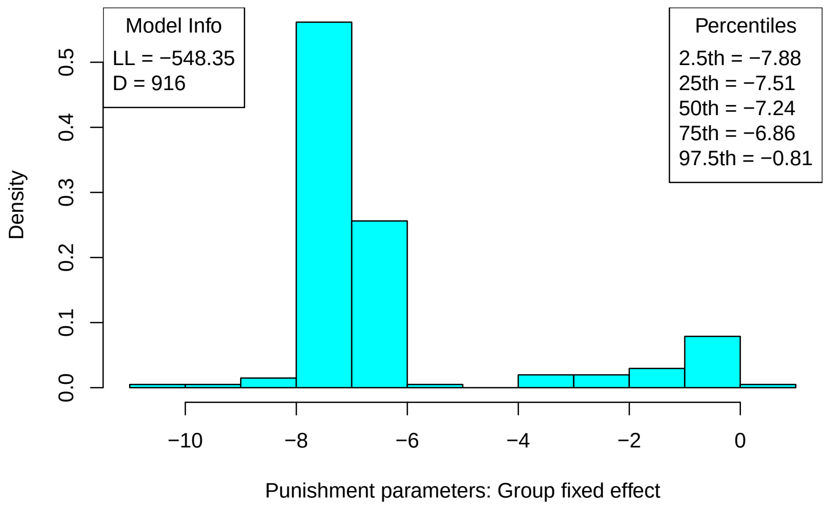
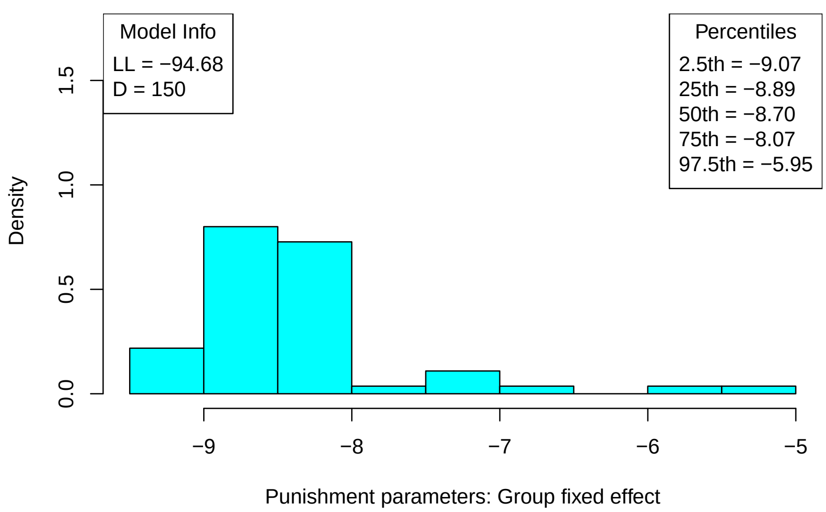


Figure E.3: Calibrated estimates of $\bar{r}$ with group fixed effects (REVMOD)



noticeable heterogeneity in the estimates. Some groups expect much harsher penalties than others. This result represents an interesting venue for future work: where do these differences comes from and what attributes of groups makes larger punishments more likely? The REVMOD data provides an interesting first cut at answer this question in the main text, but more data on non-violent cases will improve our understanding of this variance.

One potentially interesting finding from Figure E.2, is that the group that does not expect to be punished for backing down is Crimean separatists in Ukraine. The short-lived "Republic of Crimea" was a brief movement that emerged in the early 1990s after the fall of the Soviet Union, and they sought to have Crimea leave Ukraine and join the Russian Federation. The movement lasted for roughly four years before it is coded as backing down in the face of Ukrainian efforts to rein them in. The main leader of the movement was exiled to Russia, but no other lasting punishments emerged (Sasse 2007). In this case, it seems like the pro-Russian movement had little reasons to suspect punishment from the new Ukrainian state, challenged, and was mostly proved right. Likewise, the country of Georgia has small, positive punishment estimates in Figure E.1. How much of these positive benefits to backing down is driven by expectations about weak states or anticipated protection from Russia may be an interesting avenue for future work.

F Robustness to case selection

Table F.1: Robustness checks on cases

	No empires	Including coups	All uncertain cases
Constant	–2.68	–2.67	–2.60
	(0.23)	(0.18)	(0.17)
Democracy	0.09	0.06	0.12
	(0.11)	(0.08)	(0.10)
Military size	0.05	0.10	0.08
	(0.14)	(0.08)	(0.10)
Other groups	–0.21	–0.17	–0.14
	(0.15)	(0.11)	(0.13)
Separatist Dispute	0.20	0.36	0.18
	(0.19)	(0.18)	(0.19)
Log Likelihood	–856.56	–908.67	–1064.75
Observations	2272	2355	2415
χ^2	52.56	53.81	44.48
% $V_R > S_R$	100%	100%	100%
% $V_G > C_G$	100%	96%	100%
% $S_R > \bar{r}$	100%	100%	100%

Standard errors in parentheses. All non-binary covariates are standardized. Constants represent the payoff when non-binary covariates are at their average and dummy variables are 0.

In this appendix I consider how robust the main results are to the case selection. This set of checks only applies to the CONIAS data as coding these cases introduces more researcher degreesof-freedom than the REVMOD cases. The results are presented in Table F.1. In the first column, I consider a model without Empires (as defined by Fearon and Laitin 2003), as cases of colonial independence may be fundamentally different from other cases. In the second column, I include the coup cases from CONIAS, which were not included in the main model. In the third column, I consider all COINAS cases, including all cases where there was uncertainty about the coding. The point estimates on the punishment parameters are largely unchanged regardless of the case selection and across the three models the intuitive payoff restrictions are met in nearly all observations. Likewise, we see that backing down is costly (worse than the status quo) in every observation across every model.

G Robustness to repression measures

In this Appendix, I consider alternative measures of state repressive ability. In the main text, I used military personnel per capita has the main measure of repressive ability. This proxy was chosen mostly because of its good spatial and temporal coverage. We now consider the robustness of the main results when we use two common alternative measures for this concept: the political terror scale (PTS) and tax revenue divided by GDP (sometimes referred to as the tax ratio). The results are reported in Table G.1. The first two columns consider the PTS with both the CONIAS & MAR sample and the REVMOD sample. The last two columns do the same but with the tax ratio measure. Note that in both cases the number of observations decreases dramatically. Despite the reduced samples, we still find that the point estimates largely match those from the main text models.

References

Crisman-Cox, Casey and Michael Gibilisco. 2021. "Estimating Crisis Signaling Games in International Relations: Problems and Solutions." *Political Science Research and Methods* 9(3):565–582.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, insurgency, and civil war." *American Political Science Review* 97(1):75–90.

Table G.1: Alternative measures of repressive ability

	Political Terror scale		Tax revenue	
Case selection	CONIAS & MAR	REVMOD	CONIAS & MAR	REVMOD
Constant	−2.41	−0.17	−2.58	−0.38
	(0.22)	(0.34)	(0.24)	(0.33)
Political terror	−0.30	−0.03		
	(0.40)	(0.31)		
Tax revenue			0.29	0.17
			(0.27)	(0.13)
Polity	−0.08	0.14	0.23	0.11
	(0.28)	(0.36)	(0.20)	(0.22)
Other groups	−0.16	−0.46	−0.01	−0.00
	(0.41)	(0.61)	(0.24)	(0.44)
Separatist	0.07	−0.92	0.29	−1.40
	(0.31)	(0.60)	(0.30)	(0.90)
Political organization		−0.89		−0.82
		(0.52)		(0.47)
Islamic		−1.28		−1.25
		(0.68)		(0.66)
Leftist		−0.59		−0.92
		(0.59)		(0.51)
Log Likelihood	-416.62	-175.29	-556.77	-282.97
Observations	1034	340	1467	439
% $V_R > S_R$	96%	100%	47%	90%
% $V_G > C_G$	94%	22%	100%	11%
% $S_R > \bar{r}$	100%	93%	100%	100%

Notes: Standard errors in parentheses. All non-binary covariates are standardized. Constants represent the payoff when non-binary covariates are at their average and dummy variables are 0.

Jo, Jinhee. 2011. "Nonuniqueness of the Equilibrium in Lewis and Schultz's Model." *Political Analysis* 19(3):351–362.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11(4):345–367.

Sasse, Gwendolyn. 2007. *The Crimea Question: Identity, Transition, and Conflict.* Harvard University Press.