

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                |                                                      |                                                          |                                                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------------------------|------------------------------------------------------|----------------------------------------------------------|------------------------------------------------|
| <b>REPORT DOCUMENTATION PAGE</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                   |                                | Form Approved OMB NO. 0704-0188                      |                                                          |                                                |
| <p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p> |                   |                                |                                                      |                                                          |                                                |
| 1. REPORT DATE (DD-MM-YYYY)<br>14-04-2020                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                   | 2. REPORT TYPE<br>Final Report |                                                      | 3. DATES COVERED (From - To)<br>1-Jan-2016 - 31-Jan-2020 |                                                |
| 4. TITLE AND SUBTITLE<br>Final Report: The Organization that Does(n't) Bark:<br>Organizations and the Decision for Political Violence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                   |                                | 5a. CONTRACT NUMBER<br>W911NF-16-1-0023              |                                                          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 5b. GRANT NUMBER                                     |                                                          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 5c. PROGRAM ELEMENT NUMBER<br>611102                 |                                                          |                                                |
| 6. AUTHORS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                   |                                | 5d. PROJECT NUMBER                                   |                                                          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 5e. TASK NUMBER                                      |                                                          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 5f. WORK UNIT NUMBER                                 |                                                          |                                                |
| 7. PERFORMING ORGANIZATION NAMES AND ADDRESSES<br>State University of New York (SUNY) at Albany<br>Office for Sponsored Programs, MSC 312<br>1400 Washington Avenue<br>Albany, NY 12222 -0100                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                   |                                | 8. PERFORMING ORGANIZATION REPORT NUMBER             |                                                          |                                                |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)<br>U.S. Army Research Office<br>P.O. Box 12211<br>Research Triangle Park, NC 27709-2211                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                   |                                | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>ARO              |                                                          |                                                |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 11. SPONSOR/MONITOR'S REPORT NUMBER(S)<br>66997-LS.4 |                                                          |                                                |
| 12. DISTRIBUTION AVAILABILITY STATEMENT<br>Approved for public release; distribution is unlimited.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                   |                                |                                                      |                                                          |                                                |
| 13. SUPPLEMENTARY NOTES<br>The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                |                                                      |                                                          |                                                |
| 14. ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                   |                                |                                                      |                                                          |                                                |
| 15. SUBJECT TERMS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                   |                                |                                                      |                                                          |                                                |
| 16. SECURITY CLASSIFICATION OF:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                                | 17. LIMITATION OF ABSTRACT<br>UU                     | 15. NUMBER OF PAGES                                      | 19a. NAME OF RESPONSIBLE PERSON<br>Victor Asal |
| a. REPORT<br>UU                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | b. ABSTRACT<br>UU | c. THIS PAGE<br>UU             |                                                      |                                                          | 19b. TELEPHONE NUMBER<br>518-591-8729          |

# RPPR Final Report

## as of 14-Apr-2020

Agency Code:

Proposal Number: 66997LS

Agreement Number: W911NF-16-1-0023

### INVESTIGATOR(S):

**Name:** R. Karl Rethemeyer  
**Email:** kretheme@albany.edu  
**Phone Number:** 5184425283  
**Principal:** N

**Name:** Michael Young  
**Email:** michael@socialscience.net  
**Phone Number:** 6145275200  
**Principal:** N

**Name:** Victor Asal  
**Email:** vasal@albany.edu  
**Phone Number:** 5185918729  
**Principal:** Y

Organization: **State University of New York (SUNY) at Albany**

Address: Office for Sponsored Programs, MSC 312, Albany, NY 12220100

Country: USA

DUNS Number: 152652822

EIN:

**Report Date:** 30-Apr-2020

Date Received: 14-Apr-2020

**Final Report** for Period Beginning 01-Jan-2016 and Ending 31-Jan-2020

**Title:** The Organization that Does(n't) Bark: Organizations and the Decision for Political Violence

**Begin Performance Period:** 01-Jan-2016

**End Performance Period:** 31-Jan-2020

**Report Term:** 0-Other

Submitted By: Victor Asal

Email: vasal@albany.edu

Phone: (518) 591-8729

**Distribution Statement:** 1-Approved for public release; distribution is unlimited.

**STEM Degrees:** 0

**STEM Participants:** 0

**Major Goals:** Our proposed research program develops a theoretical framework for understanding the use of violence by non-state organizations and organizational decisions to escalate, de-escalate, or vacillate in their use of violence (Gurr, 1970; Rasler 1996). We focus on how organizational decisions are influenced both by inter-organizational dynamics and by government retaliatory and consolatory actions towards non-state groups within a polity. In particular, we argue that (a) pre-existing historical group ideological characteristics (religious, cultural, and political), (b) domestic and international political opportunity structures, and (c) intergroup coalition networks condition the initial decision of a non-state organization to use violence (Tilly, 1973; Lichbach, 1994; Gates, 2002; Berman and Laitin, 2008; Horowitz, 2010). Certain government retaliatory actions to domestic groups connected to the organization, like publically imprisoning nonviolent activists, make a change from nonviolent to violent strategies more likely as these actions undermine trust in the government and credibility in government conciliatory gestures. Conversely, international organizational contacts and the government's expansion of association rights make a move towards nonviolent tactics more likely for non-state groups with a history of violence (Pierskalla, 2010, Siegel, 2011 Bell et al., 2012).

Using these theoretical priors, we will use a novel quantitative approach that allows us to simultaneously examine the ideology, behavior, and position of a non-state actor within both its organizational and organizational-country network. One of the key challenges when using quantitative methods to examine social and political stability is that there is very little data that captures non-state actor organizational factors (such as ideology, size, network connections, etc.) that may be related to movements or organizations that are challenging or supporting governments. To address this issue we will build on work done with one of the few existing datasets that captures the organizational data on both violent and nonviolent groups across the Middle East region – The Minorities at Risk Organizational Database (MAROB). These data allow us to identify key factors that help to explain why organizations (a) begin, continue, or end the use of violence and/or protest, (b) choose a mixed strategy that

## RPPR Final Report as of 14-Apr-2020

involves both violence and non-violence, and (c) the impact of government behavior. These factors are both contextual (state-based, culturally-based, regionally-based) and organizational (size, ideology, religion, leadership traits and the sentiments about the group in the society).

Previous MAROB research (as well as research using the related Political Organization Database that codes a sample of Middle Eastern Extremist organizations both violent and not) has used standard logistic analysis to analyze the decision to use violence as well as the decision to use different types of contentious political behavior from protest to violence (Asal et al. 2013, Asal et al. forthcoming1, Asal et al. forthcoming 2). This analysis has already shed new light on the choice of violence within the context of ethnopolitical organizations in the Middle East. For example, one analysis found that religious ideology is not impacting the decision to use violence as much as it is impacting the decision to target civilians once violence is chosen as a strategy (Asal et al. forthcoming 1). Another analysis underlined the importance that gender ideology can have on the impact of the choice of violence; in that organizations with such an ideology were much less likely to choose violence (Asal et al. 2013). Previous MAROB analysis also found that government behavior had a big impact on how organizations chose to behave in terms of the kinds of contentious tactics they chose (Asal et al. 2013, Asal et al. forthcoming1, Asal et al. forthcoming2). The fact that the MAROB data includes comparable violent and nonviolent organizations has also allowed for the ability to identify if a group's participation in violence has an impact on organizational splits (Asal et al. 2012a), as well as the propensity of an organization to be involved in the illegal drug trade (Asal et al. 2012b), or the likelihood that they will engage in transnational activity (Asal et al. forthcoming 3).

However, despite these important results, the existing MAROB data and previous analyses of political violence have some important limitations that prevent definitive testing of many questions:

- **Limited theoretical scope.** Much of the existing research on terrorism focuses only groups that use violence and the violence they use (see, for example, Choi, Seung-Whan, 2010). This creates a problem for analysis in that null cases (groups that do not use violence) are missing and analysis may reveal effects that appear real but are in fact spurious. A full test of the causal importance of the identified factors requires the presence of a substantial number of null cases.
- **Data Analysis.** Most analysis of political violence within the extant literature employs some form of logistic analysis to model the forces that shape whether an organization is engaged in violence or not. Although such an approach provides useful information about factors connected to political violence, it only provides a narrow snapshot view of violence. In this snapshot, groups are either violent or they are not. A richer analysis would not only seek to understand the determinants of whether a group is violent or not, but also what influences how long the group continues to use violence as a strategy for achieving its goals.
- **Limited data.** The original MAROB data set runs from 1980 to 2004 and covers 118 ethno-political organizations in 16 countries of the Middle East and North Africa. To fully test theories that explain the use and non-use of violence the dataset should be extended in time to the present day and extended in scope to cover various types of political groups (including non-violent, non-contentious groups) around the entire globe. Unfortunately, as one unsuccessful attempt to extend MAROB demonstrates, replicating the original effort is extremely difficult, time consuming and costly.
- **Missing data.** The existing MAROB dataset does not contain network data beyond simple dyads, and our existing terrorist datasets only allow us to examine the relational dynamics of violent behavior. We have no window into the relational dynamics of non-violent behavior and transitional behavior (that is, the decision to switch from contentious to violent political means). For example, does rivalry drive organizations to choose violent behaviors, much as rivalry drives organizations already engaged in violence to engage in more extreme violence? Or does rivalry in some cases drive organizations to differentiate – that is, to choose nonviolent means so that the focal group serves as an alternative for those members of a minority group that wish to express dissatisfaction with government action but not through violent means? Is violence “contagious”: if one organization representing an at-risk population chooses violence, does that draw all others into violence through outbidding? We know from analysis of the existing data that organizations transition back and forth between periods of violence and nonviolence; network relations may provide a missing explanatory piece. Fundamentally, we wish to examine the degree to which network and behaviors co-evolve across time using stochastic social network methods, but to pursue this line of analysis we need longitudinal data on alliances and rivalries among MAROB groups. MAROB currently treats each organization it covers in isolation as though it has no interaction with other groups.

This effort will address each of these limitations.

# RPPR Final Report

as of 14-Apr-2020

**Accomplishments:** See upload section

**Training Opportunities:** Research assistants were trained in coding scheme development and worked on the coding schemes.

**Results Dissemination:** Nothing to Report

**Honors and Awards:** Nothing to Report

**Protocol Activity Status:**

**Technology Transfer:** Nothing to Report

## PARTICIPANTS:

**Participant Type:** PD/PI

**Participant:** Victor Asal

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

**Participant Type:** Consultant

**Participant:** Amanda Murdie

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

**Participant Type:** Consultant

**Participant:** James Michael Greig

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

**Participant Type:** Consultant

**Participant:** Joseph Young

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

**Participant Type:** Consultant

**Participant:** Michael Young

**RPPR Final Report**  
as of 14-Apr-2020

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

**Participant Type:** Other Professional

**Participant:** Suzanne Weedon Levy

**Person Months Worked:** 1.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

**Funding Support:**

***The Organization that Does(n't) Bark: Organizations and the  
Decision for Political Violence***

Army Research Office

W911NF-16-0023

## **Table of Contents**

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Introduction.....                                               | 3   |
| BARK Data Coding Collection .....                               | 6   |
| Hazard Modeling Analysis .....                                  | 21  |
| Causal Pathway Analysis.....                                    | 39  |
| Co-Evolution Modeling .....                                     | 58  |
| Appendix I: The Organization that Does(n't) BARK Codebook ..... | 71  |
| Appendix II: BARK and MAROB Coding Comparison .....             | 101 |
| Appendix III Original Proposal .....                            | 119 |

## Introduction

The goal of the project “The Organization that Does(n’t) Bark: Organizations and the Decision for Political Violence” was to address the lack of “null cases” in so much of the quantitative research on the start and use of violence by organizations when analyzing the use of violence and to bring automated coding much more fully into efforts to try and understand why some organizations turn to violence and others do not. We built on the Minorities at Risk Organizational Database (MAROB) dataset, one of the only datasets that collected yearly data on both violent and nonviolent as well as non-contentious political organizations, and we created a team to collect automated coded data. This data was analyzed to better understand why organizations start being contentious, why they start using violence and why they stop. Once the data was collected, we applied a variety of forecasting models and hazard models – some of which have been rarely used in political science efforts on this topic. Data was recorded for violent and non-violent organizations in the Middle East and North Africa (MENA) as well as Eastern Europe. This report contains four sub-reports and three appendices. The four sub-reports are:

- Bark Data Coding Collection
- Hazard Modeling Analysis
- Causal Pathway Analysis
- Co-Evolution Modeling

The three appendixes are:

- The detailed codebook that was used to collect the data.
- A comparison of the data collected in the automated coding effort in this project with the human data collected in the original MAROB effort.
- The original proposal that was the basis for this research effort and contains an in-depth look at the motivation for the research in this effort.

The Bark Data Coding Collection report by Michael D. Young and Michelle L. Young lays out the process by which the project variables were prioritized and coded. It also lays out the analysis procedure for measuring the accuracy of the coded variables and the approach to data collection and open source research used. We improved the quality of the data throughout the project both in terms of recall and accuracy. The recall goal of  $\geq 0.9$  was achieved for five (5) MAROB variables although the accuracy goal of  $\geq 0.6$  was achieved only for one (1) of those following the decision to increase the number of variables available for analysis. The effort also demonstrated automated coding is cheaper and able to process many more documents than human hand coding. It can also be done closer to real time than human coding.

The Hazard Modeling Analysis by J. Michael Greig and Joseph K. Young focuses on using the generated data to examine the factors that lead organizations to become involved in violent conflict, protest, or elections using hazard modeling analysis. In this analysis, the focus is on the characteristics of a non-state organization, such as their ideology, ability to control territory, and the structural conditions that are present in an organization’s operating

environment. In addition, the government responses to organizations are also considered, as is the prior behavior of the organization. Finally, the analysis also accounts for the activity of other non-state actors in the same state and how that might impact the behavior of the organization in question. The hazard model is used to examine the initiation of certain behaviors by the organization. The analysis finds that the initiation of protests, violence, and negotiations tend to be driven by distinct casual processes. In general, the initiation of protests, as opposed to the initiation of violence, tends to be more directly driven by the organizational characteristics of non-state organizations. In contrast, the initiation of violence by non-state organizations is more connected to the structural conditions in which that organization operates than to the onset of protests or negotiations. However, there are commonalities in factors that impact initiation of both protests and violence. The way in which a non-state organization's government responds to its grievances plays an important role in shaping the likelihood the organization will engage in protests, or negotiations, or adopt violence. They also find evidence for an important set of cascade effects in which the actions of other non-state organizations within a state influence the likelihood that a non-state organization will undertake protests, violence, or negotiations.

The Causal Pathway Analysis by Amanda Murdie also looks at the choice of violence – and specifically the use of terrorism, measured by the number of terrorist incidents in a year. A causal pathway approach, which allows for the quantitative assessment not only of the *direct* effect of certain characteristics on non-state organizational violence but the *indirect* effects of these characteristics on violence through unique intermediary steps, is used that looks at the indirect effects of the different characteristics of the organization on this choice. This report finds that the key factor opening a pathway towards violence is the ineffectiveness of the home government. This impact is mediated by state repression of the organization, the size of the organization, and an organization's ability to control territory. . The analysis done using all the BARK data finds strong support for the impact of government ineffectiveness on the use of terrorism by the organizations in the dataset. For the region of Eastern Europe, the results are similar but the relationship between government ineffectiveness and the mediators, is not statistically significant. For the Middle East and North Africa region, the government ineffectiveness measure generally fails to meet conventional levels of statistical significance. The hypothesized mediators do affect organizational violence. Using these causal pathway models, the analysis finds that government ineffectiveness leads to a greater likelihood of territorial control by the organization and the increased use of violent repression by the government. These factors ultimately spur more organizational violence. In sum, by focusing on improving government effectiveness, the process may be disrupted, limiting the direct and indirect pathways driving organizational violence.

The Co-Evolution Modeling by R. Karl Rethemeyer and Victor Asal looked specifically at the violent organization's lethality and Violent Nonstate Actors (VNSA) network/behavior *co-evolution*. The analysis found a strong tendency for organizations to *avoid* making connections. However, the modeling does find three strong bases for connections linked to homophily: shared ethnonationalist ideology, shared religious ideology, and/or shared separatist ideology. Territorial control is an important factor that drives organizations to make alliances. The same is true for size, large organizations tend to have more connections and to prefer to have connections to organizations of a similar size. The analysis also suggests an accelerator effect on killing from the actions of peers: if an organization's allies are killing a lot, its level of killing will also go up, but this increase is dependent on amount of killing by that organization's allies.

In sum, this effort has made considerable progress in generating automated data for the analysis of why some organizations embrace violence and others do not. Overall, our current analysis found that automated data shined a light on mobilization efforts by organizations as well as the choice of violence and targets. Key factors are the characteristics of an organization such as ideology and size, in addition to the behavior of their allies. These characteristics drive contentious behaviors, as well as the likelihood of forming network connections with potential allies. The structural environment of the organization also has an important impact on its behavior. Specifically, the capability of the organization's homebase state also plays a role in the behavior of that organization. Non-state actors in countries with ineffective governments are more likely to control territory and be subject to repressive state policies. We believe that these findings are useful theoretically and have important policy implications as well. We hope to extend these efforts in the future.

## ***BARK Data Coding Collection***

Michael D. Young & Michelle L. Young  
Social Science Automation  
[michael@socialscience.net](mailto:michael@socialscience.net)

The goals established for BARK data collection were 1) to develop and then refine automated coding schemes for at least five (5) of the MAROB variables, 2) to acquire documents from Lexis-Nexis news aggregation service from a wide range of sources for the years 2005-2017, and 3) to code the acquired documents and generate data for all MAROB organizations for the years from 2005-2017.

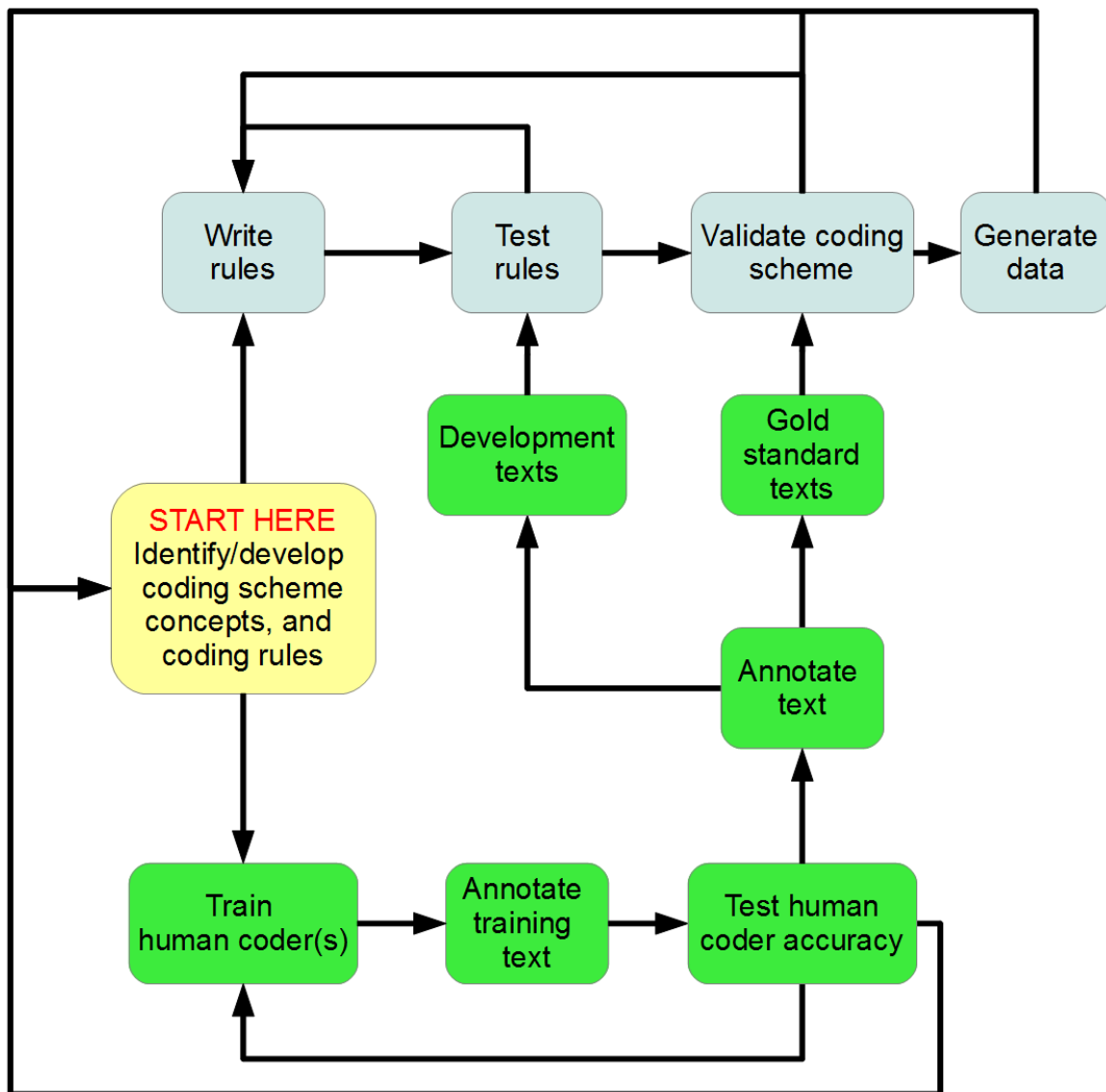
## Coding Scheme Development

The initial priority ordering of thirty-four (34) MAROB variables was:

- |                        |                                                |
|------------------------|------------------------------------------------|
| 1. Size                | 20. Allies&Rivals                              |
| 2. Territorial control | 21. Organizational Violence<br>(against state) |
| 3. Democratic          | 22. Violent Rhetoric (Target<br>Levels)        |
| 4. Environmental       | 23. INGO                                       |
| 5. Ethnic              | 24. IGO                                        |
| 6. Leftist             | 25. Hold office                                |
| 7. Nationalist         | 26. Run for Office                             |
| 8. Regime_Change       | 27. Fd.Drug                                    |
| 9. Religious           | 28. Fd.Smuggling                               |
| 10. Rightist           | 29. Fd.State Sponsor                           |
| 11. Secular            | 30. Organizational Negotiations                |
| 12. Separatist         | 31. Protest Location                           |
| 13. Status-Quo         | 32. Inter Org conflict<br>(Interorgcon)        |
| 14. Supremacist        | 33. ORGReb                                     |
| 15. State Repression   | 34. Leadership                                 |
| 16. Protest            |                                                |
| 17. State Violence     |                                                |
| 18. Mass Killing       |                                                |
| 19. Suicide            |                                                |

Following an evaluation of the variables and their expression in source texts, the Size, Territory, and twelve (12) Ideology variables (Democratic, Environmental, Ethnic, Leftist, Nationalist, Regime\_Change, Religious, Rightist, Secular, Separatist, Status-Quo, Supremacist) were selected for automation with addition variables added during the period of performance as resources allowed (Violent repression, Sexual violence, Drug Funding, Protest) or research priorities indicated (Allies&Rivals). In addition, Recall (  $\frac{[True\ Positive]}{[True\ Positives + False\ Negatives]}$  ) was established the primary criteria for coding scheme performance because: 1) information is sparse in the source documents, 2) relatively few quality-controlled exemplars were available for the variables and 3) the planned data production process included human review of each observation. This set of factors made false positives less problematic than false negatives. The desired value for Recall was set at  $\geq 0.9$  with a secondary goal of Accuracy (  $\frac{[True\ Positives]}{[True\ Positives + False\ Positives + False\ Negatives]}$  ) set at  $\geq 0.6$ . True Negatives are not included in the Accuracy calculation because doing so gives a value misleadingly close to 1.0. As the project progressed, the research team elected to increase the number of variables coded to allow for broader analysis rather than focus on meeting both criteria for each variable before developing additional coding schemes.

Ideally, coding scheme development (see figure X below from Young and Hermann, 2016), would draw on a well-developed concept/annotation guide and a large set of annotated exemplars for each possible coding. For this project, these resources did not exist so the first step for each of the target variables was the creation of a minimal annotation guide and annotated articles by human coders following an adapted version of the original coding process but using small sections of articles known to contain examples of evidence used for each particular variable.



In parallel to the creation of the annotated articles, a base coding scheme stack was created and the existing MAROB organization name and alias lists were added to the entity identification coding scheme resulting in a base stack used to perform functions for all of the variable specific

coding schemes (see table below). The use of this base stack for multiple variables increases the efficiency of the coding process while still allowing “destructive” changes to the text for each variable.

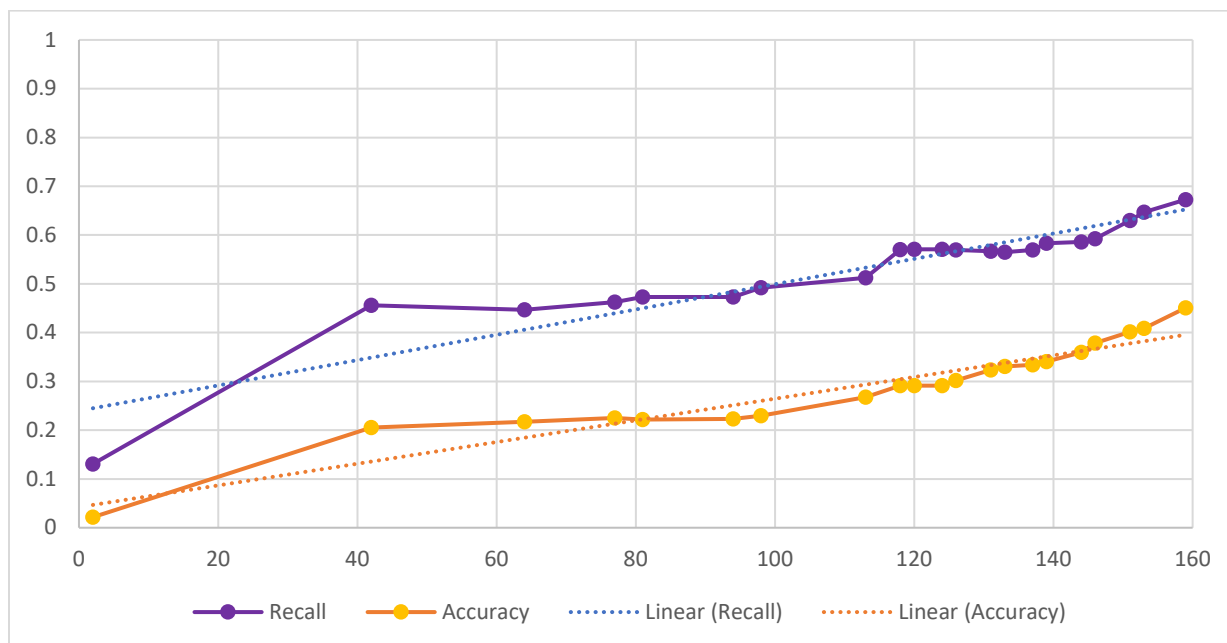
| <b>Table/Subscheme</b>                    | <b>Function</b>                                                                                  |
|-------------------------------------------|--------------------------------------------------------------------------------------------------|
| POST                                      | Part of speech tagging.                                                                          |
| GoodBad                                   | Identify positive and negative terms.                                                            |
| HumanMaleFemale                           | Identify names and references to humans.                                                         |
| Entity_Recognition                        | Identify named entities, including organizations of interest.                                    |
| VerbCollapse                              | Combine auxiliary verbs and tense markers with main verb of verb phrase.                         |
| Org_cues                                  | Identify phrases that may indicate the presence of an organization of interest.                  |
| Identify_Org                              | Reconcile org_cues with identified organizations if possible.                                    |
| Different_Org_same_name_different_country | Not implemented.                                                                                 |
| Org_anaphora                              | Reconcile organization anaphora (group, militia, etc) with identified organizations if possible. |

For example, the Allies&Rivals scheme builds on the base stack by adding:

| <b>Table</b>                  | <b>Function</b>                                                                                                  |
|-------------------------------|------------------------------------------------------------------------------------------------------------------|
| Tag_Allies_and_Rivals         | Identify terms and phrases that identify conflict (rivals) or support (allies) between two or more groups.       |
| Remove_False_Positive1        | Remove false positives (terms that should not be coded).                                                         |
| Remove_False_Positive         | Remove false positives (terms that should not be coded).                                                         |
| Make_Org_List                 | When no Allies&Rivals terms is present in a sentence, create a list of organizations of interest in the sentence |
| Special Case Rules            | Additional rules for complicated contexts                                                                        |
| Consolidate_Allies_and_Rivals | Associate organizations with Allies&Rivals terms                                                                 |
| Remove_Orgs                   | Delete all organization references not associated with an Allies&Rivals term.                                    |
| ARO production ARO output     | Generate output.                                                                                                 |

The first step, Tag\_Allies\_and Rivals, Identifies terms and phrases that identify conflict (rivals) or support (allies) between two or more groups and marks them for future references as an Allies& Rivals term and either ally or rival specific. For example, the word “support” would have an allies modifier while the word “conflict” would have a rivals modifier. This process involves a pass through the document in which terms are labeled Allies&Rivals based upon their inclusion in the dictionaries created for this scheme.

Once the tagging rule set scans through and tags all of the words and phrases, there are two rule sets that remove false positives, or terms that should not be coded. These rule sets focus on context clues within the sentence to determine if the material is codeable. If it is not codeable information, then the Allies&Rivals tag is removed. For example, if the word “conflict” is in a sentence it would be tagged in the initial rule set, however if there are no organizations linked to the word “conflict” then it should not be coded. After the remove false positive phase is complete, the Make\_Org\_List rule set is used to create a list of organizations in all sentences without an Allies&Rivals marker for potential use in conjunction with the preceding and following sentences. For example, if the sentence “These organizations are in conflict with Group A” is referring to organizations from the previous sentence, then those organizations must somehow be associated with the term “conflict”. The next rule set, “Special\_Case\_Rules focuses on more contextually complicated issue whether false positive or false negative. Once the identification of terms and the removal of false positives is complete and only the codeable information is left, the Consolidate\_Allies\_and\_Rivals rule set connects organizations to the codeable term. The other coding schemes all have similar structures although the lexicons and contexts vary.



The level of effort required to develop each coding scheme (one or more variables) varies depending on the complexity of the coding, the availability of appropriate sub-schemes, and the experience of the coding scheme developer. Allies&Rivals was the last coding scheme developed by an experienced undergraduate research assistant who previously worked on the Ideology coding scheme, perhaps the best case scenario. Nonetheless, the pattern of improvement and the level of effort needed to achieve the recall and accuracy goals is representative of the other development efforts; 200 hours of effort by a trained undergraduate research assistant is a good baseline for estimating future projects.

The final recall and accuracy values for the coding schemes are shown in the table below. The recall goal of  $\geq 0.9$  was achieved for five (5) MAROB variables although the accuracy goal of  $\geq 0.6$  was achieved only for one (1) of those following the decision to increase the number of variables available for analysis.

| Coding Scheme      | # of MAROB Variables | Recall (in-sample) | Accuracy (in-sample) |
|--------------------|----------------------|--------------------|----------------------|
| Size               | 1                    | 0.98               | 0.82                 |
| Territory control  | 1                    | 0.93               | 0.58                 |
| Ideology*          | 12                   | 0.72               | 0.50                 |
| Violent repression | 1                    | 0.98               | 0.11                 |
| Sexual violence    | 1                    | 0.92               | 0.36                 |
| Drug Funding       | 1                    | 0.85               | 0.57                 |
| Allies & Rivals    | 1                    | 0.67               | 0.45                 |
| Protest**          | 1                    | 0.98               | 0.14                 |
| Find Orgs          | -                    | 0.97               | 0.41                 |

\* Ideology includes 12 of the original MAROB variables: Democratic, Environmental, Ethnic, Leftist, Nationalist, Regime\_Change, Religious, Rightist, Secular, Separatist, Status-Quo, Supremacist.

\*\*Protest data was not used for analysis due to issues using the coding scheme in the production environment.

## Document Acquisition

Documents were acquired from Lexis-Nexis via a custom C# program using the (now-discontinued) Web Services Kit. The C# program used a reference table of sources, start dates and number of days for a single search to establish search parameters. The research access purchased for BARK (10pm Friday to 6am Monday) permitted:

- 2000 documents per search
- 750 searches per hour
- 300,000 documents retrieved per day.

Up to four (4) downloads were running at a time with manual scripts run and inspection undertaken to verify all the documents have been downloaded for a given source. The retrieved documents are stored by source and then by Year-month, and the filename is the document “Globally Unique Identifier” from the original source provider. This was done for two reasons: 1) to easily select documents with a specific date range, and 2) to avoid file system limitations on the number of documents in a directory. Although most projects do not encounter this limitation, BARK required the storage of millions of documents

## Data Coding

For the data coding process another custom C# program is used to transform each of the documents from their original XML format, to an appropriate input format for the coding engine (Profiler Plus) and copy the files to the servers used to code the documents and to create. An existing program, ConfigCreator was used to generate PPB (driver) files for Profiler Plus.

Finally, on each data production server a program, the Controller, sends out assignments to the Profiler Plus clients. Each set of source documents specified in the PPB files is run through the schemes created by coders wherein the program runs through and grabs results from each of the documents. The optimal number of Profiler Plus clients for a server is  $n-2$ , where  $n$  is the number of cores in the server processor(s). For BARK, there were three controllers on two servers with 3-7 clients each maximizing efficiency by keeping coding schemes and partially processed (base stack) documents in memory from run to run and scheme to scheme.

After each file is processed a result file is created even if no data is generated. Although this increases input/output activity, it provides the ability to rerun and replace the results if errors were encountered or a coding scheme was updated. The Profiler Controller “reads” the results of the Profiler run and determines if it was successful or not. Successful runs create a ppoutput.xml file indicating where the output files(s) for the run are and place it in a directory for the output loader. Unsuccessful runs also create a similar file, but with additional details of the errors found, and placed in an error directory for manual handling, along with the corresponding PPB.

A C# loader program reads the ppoutput.xml files to find the locations of the output files, reads the output files, and loads the data into a SQL Database. Depending on the scheme output and what is required for the reporting, there may be some post processing to summarize or transform the data further as part of the load. The final Excel data submitted to the analysis team is created from SQL Stored Procedures that summarize the results in the database. The data generation process is almost all automated and requires only minimal manual effort to identify problems, fix holes, and reconfigure the process.

The table below lists the sources selected for analysis and the number and date range of documents retrieved and processed. The total number of documents processed was 9,370,951.

| Source | SourceName                                          | Files Downloaded | Files With Results | Results | Pct Files results | Pct tot files | Pct total results | timeframe covered |
|--------|-----------------------------------------------------|------------------|--------------------|---------|-------------------|---------------|-------------------|-------------------|
| AFP    | Agence France Presse - English                      | 2120916          | 160611             | 325084  | 7.6%              | 29.7%         | 34.1%             | all               |
| APINTL | Associated Press International                      | 1366324          | 78809              | 148387  | 5.8%              | 14.6%         | 15.6%             | all               |
| AFRNWS | Africa News                                         | 2868753          | 48376              | 75712   | 1.7%              | 8.9%          | 8.0%              | all               |
| AFRWEN | AllAfrica Newspapers (English)                      | 1270295          | 40434              | 61277   | 3.2%              | 7.5%          | 6.4%              | 2008+             |
| APONL  | Associated Press Online                             | 464511           | 25026              | 46699   | 5.4%              | 4.6%          | 4.9%              | 2005-2016         |
| DNSL   | Daily News (Sri Lanka)                              | 201866           | 19602              | 24735   | 9.7%              | 3.6%          | 2.6%              | 2008+             |
| ALJZEE | Al Jazeera - English                                | 46665            | 10761              | 25387   | 23.1%             | 2.0%          | 2.7%              | 2014+             |
| ASSAM  | Assam Tribune (India)                               | 107353           | 10255              | 12990   | 9.6%              | 1.9%          | 1.4%              | 2012+             |
| ARABIY | AlArabiya.net                                       | 46686            | 10209              | 22078   | 21.9%             | 1.9%          | 2.3%              | 2009+             |
| CHIDLY | China Daily                                         | 339372           | 8194               | 13804   | 2.4%              | 1.5%          | 1.5%              | 2009+             |
| HERSUN | Herald Sun/Sunday Herald Sun (Melbourne, Australia) | 1047582          | 6783               | 9779    | 0.6%              | 1.3%          | 1.0%              | all               |
| CAPTIM | Cape Times (South Africa)                           | 250392           | 6529               | 10495   | 2.6%              | 1.2%          | 1.1%              | 2007+             |
| BALTNS | Baltic News Service                                 | 320253           | 6464               | 9260    | 2.0%              | 1.2%          | 1.0%              | all               |
| CHDLY  | Chicago Daily Herald                                | 339372           | 6398               | 9606    | 1.9%              | 1.2%          | 1.0%              | 2005-8,2013-16    |
| ABCNEW | ABC News Transcripts                                | 192251           | 6289               | 8860    | 3.3%              | 1.2%          | 0.9%              | all               |
| AVOICE | Afternoon Voice (India)                             | 54934            | 6241               | 8461    | 11.4%             | 1.2%          | 0.9%              | 2012+             |
| ANEINT | Asian News International (ANI)                      | 136798           | 5264               | 7970    | 3.8%              | 1.0%          | 0.8%              | 2009-2011         |
| RECORD | Daily Record and Sunday Mail                        | 985397           | 4636               | 6367    | 0.5%              | 0.9%          | 0.7%              | 2005-2016         |
| DLYNTN | Daily Nation (Kenya)                                | 125100           | 4309               | 6959    | 3.4%              | 0.8%          | 0.7%              | 2009+             |
| ARBUNW | Arab News                                           | 61945            | 3961               | 7371    | 6.4%              | 0.7%          | 0.8%              | 2009+             |
| BGRDLY | Bangor Daily News (Maine)                           | 368429           | 3434               | 4296    | 0.9%              | 0.6%          | 0.5%              | all               |
| BALEXP | Daily Balochistan Express                           | 93642            | 3297               | 5003    | 3.5%              | 0.6%          | 0.5%              | 2010+             |
| ALAKEN | Al-Akhbar English                                   | 9136             | 3287               | 8163    | 36.0%             | 0.6%          | 0.9%              | 2012-2014         |
| DSTAR  | Daily Star                                          | 852174           | 3185               | 3886    | 0.4%              | 0.6%          | 0.4%              | 2005-2016         |
| BAHRNA | Bahrain News Agency (BNA)                           | 67323            | 2999               | 3867    | 4.5%              | 0.6%          | 0.4%              | 2009+             |
| AFRDIL | Daily Independent (Lagos)                           | 71538            | 2984               | 4590    | 4.2%              | 0.6%          | 0.5%              | 2010-2016         |
| AGEAUS | The Age (Melbourne, Australia) - Online             | 161123           | 2939               | 4897    | 1.8%              | 0.5%          | 0.5%              | 2013+             |

|        |                                               |        |      |      |       |      |      |           |
|--------|-----------------------------------------------|--------|------|------|-------|------|------|-----------|
| STJREG | The State Journal-Register (Springfield, IL)  | 245336 | 2813 | 3657 | 1.1%  | 0.5% | 0.4% | all       |
| BIRPST | Birmingham Post                               | 282950 | 2692 | 3826 | 1.0%  | 0.5% | 0.4% | x         |
| ASIAMT | South Asia Monitor                            | 10745  | 2566 | 4355 | 23.9% | 0.5% | 0.5% | 2015+     |
| TPRESS | The Press (Christchurch, New Zealand)         | 377213 | 2512 | 3646 | 0.7%  | 0.5% | 0.4% | all       |
| EVECHR | Evening Chronicle (Newcastle,UK)              | 591058 | 2484 | 3025 | 0.4%  | 0.5% | 0.3% | all       |
| DNHT   | Daily National Herald Tribune                 | 18125  | 2465 | 4152 | 13.6% | 0.5% | 0.4% | 2012+     |
| DNWSSA | Daily News (South Africa)                     | 162180 | 2440 | 3675 | 1.5%  | 0.5% | 0.4% | 2006+     |
| INDNIG | Daily Independent (Nigeria)                   | 54057  | 2244 | 3339 | 4.2%  | 0.4% | 0.4% | 2015+     |
| PNNENG | Palestinian News Network (PNN) - English      | 17666  | 2142 | 3840 | 12.1% | 0.4% | 0.4% | 2009+     |
| AIP    | Afghan Islamic Press                          | 7099   | 2140 | 2868 | 30.1% | 0.4% | 0.3% | 2011-2013 |
| COLMBN | The Columbian (Vancouver, WA)                 | 118753 | 2081 | 3289 | 1.8%  | 0.4% | 0.3% | all       |
| CAPITL | The Capital (Annapolis, MD)                   | 169157 | 1925 | 2852 | 1.1%  | 0.4% | 0.3% | 2005-2014 |
| ARABEN | Al-Arab (English)                             | 6695   | 1857 | 4654 | 27.7% | 0.3% | 0.5% | 2012-2015 |
| WAIKTO | Waikato Times (Hamilton, New Zealand)         | 238806 | 1776 | 2729 | 0.7%  | 0.3% | 0.3% | all       |
| CHIDAW | China Daily - Africa Weekly                   | 159706 | 1668 | 2457 | 1.0%  | 0.3% | 0.3% | 2013=     |
| QTRNEN | Qatar News Agency (QNA) (English)             | 40961  | 1400 | 2115 | 3.4%  | 0.3% | 0.2% | 2013+     |
| CAPTMS | The Capital Times (Madison, Wisconsin)        | 91117  | 1301 | 1678 | 1.4%  | 0.2% | 0.2% | all       |
| AFRNEW | Africa Newswire                               | 24399  | 1300 | 2328 | 5.3%  | 0.2% | 0.2% | 2016+     |
| CHIDUS | China Daily - US Edition                      | 98345  | 1113 | 1617 | 1.1%  | 0.2% | 0.2% | 2014+     |
| BANEGA | Bahrain News Gazette                          | 22383  | 991  | 1372 | 4.4%  | 0.2% | 0.1% | 2016+     |
| CALSUN | Calgary Sun                                   | 93636  | 981  | 1433 | 1.0%  | 0.2% | 0.2% | 2016+     |
| CHIDHK | China Daily (Hong Kong Edition)               | 25233  | 919  | 1369 | 3.6%  | 0.2% | 0.1% | 2013+     |
| APO    | African Press Organization                    | 22722  | 830  | 1168 | 3.7%  | 0.2% | 0.1% | 2013+     |
| DADISA | Daily Dispatch (South Africa)                 | 60754  | 786  | 1078 | 1.3%  | 0.1% | 0.1% | 2014+     |
| ALGPSV | Algeria Press Service                         | 31395  | 766  | 997  | 2.4%  | 0.1% | 0.1% | 2009+     |
| SSTAR  | The Sunday Star-Times (Auckland, New Zealand) | 87526  | 732  | 1073 | 0.8%  | 0.1% | 0.1% | all       |
| AFRDCL | Daily Champion (Lagos)                        | 16445  | 547  | 839  | 3.3%  | 0.1% | 0.1% | 2010-2012 |
| DLNWSA | Daily News(S. Africa)                         | 35127  | 537  | 772  | 1.5%  | 0.1% | 0.1% | 2011+     |
| BIHAR  | Bihar Times (India)                           | 7957   | 495  | 588  | 6.2%  | 0.1% | 0.1% | 2011+     |
| USNEWS | U.S. News & World Report                      | 7540   | 435  | 833  | 5.8%  | 0.1% | 0.1% | 2005-2010 |

|         |                                                 |       |     |     |       |      |      |                           |
|---------|-------------------------------------------------|-------|-----|-----|-------|------|------|---------------------------|
| DLYREC  | The Daily Record (Baltimore, MD)                | 48869 | 389 | 480 | 0.8%  | 0.1% | 0.1% | all                       |
| AFRTIF  | The Independent (Kampala, Uganda)               | 7723  | 384 | 723 | 5.0%  | 0.1% | 0.1% | 2010+                     |
| AFRCCTE | Cameroon Tribune (Yaounde) English              | 12286 | 366 | 584 | 3.0%  | 0.1% | 0.1% | 2010+                     |
| AFRCCT  | Concord Times (Freetown)                        | 9687  | 301 | 378 | 3.1%  | 0.1% | 0.0% | 2010+                     |
| CAMBGA  | Cambodia News Gazette                           | 7787  | 246 | 412 | 3.2%  | 0.0% | 0.0% | 2016+                     |
| AZNEGA  | Azerbaijan News Gazette                         | 17266 | 214 | 344 | 1.2%  | 0.0% | 0.0% | 2016+                     |
| ALJAZR  | Al Jazeera Africa                               | 1661  | 175 | 356 | 10.5% | 0.0% | 0.0% | 2016+                     |
| BOTDAN  | Botswana Daily News(Asianet-Pakistan (Pvt) Ltd) | 18636 | 170 | 184 | 0.9%  | 0.0% | 0.0% | 2014+                     |
| BRUGAZ  | Brunei News Gazette                             | 10137 | 113 | 171 | 1.1%  | 0.0% | 0.0% | 2016+                     |
| ASEANT  | ASEAN Tribune                                   | 6117  | 87  | 127 | 1.4%  | 0.0% | 0.0% | 2016+                     |
| BTIMES  | Bali Times (Indonesia)                          | 1408  | 72  | 92  | 5.1%  | 0.0% | 0.0% | 2012-2016                 |
| CREST   | Crest                                           | 2004  | 63  | 107 | 3.1%  | 0.0% | 0.0% | 2010-2013                 |
| CENQUN  | Central Queensland News                         | 35503 | 51  | 57  | 0.1%  | 0.0% | 0.0% | 2009+                     |
| BHACHR  | Bharat Chronicle                                | 530   | 39  | 53  | 7.4%  | 0.0% | 0.0% | 2008-2009                 |
| MBUSTM  | Business Times (Malaysia)                       | 11535 | 31  | 38  | 0.3%  | 0.0% | 0.0% | 2005-14<br>very<br>sparse |
| AFRAIM  | Agencia de Informacao de Mocambique (English)   | 11221 | 28  | 35  | 0.2%  | 0.0% | 0.0% | 2010+                     |
| CALOUN  | Caloundra Weekly (Queensland)                   | 25143 | 23  | 28  | 0.1%  | 0.0% | 0.0% | 2009+                     |
| BOBSER  | Bhutan Observer                                 | 1165  | 23  | 26  | 2.0%  | 0.0% | 0.0% | 2012-2015                 |
| DLPKST  | Daily Pakistan Today                            | 96278 | 22  | 42  | 0.0%  | 0.0% | 0.0% | 2011+                     |
| AFRMAN  | African Manager (English)                       | 6824  | 20  | 31  | 0.3%  | 0.0% | 0.0% | 2010+                     |
| BUSTEL  | Business Telegraph                              | 16345 | 18  | 20  | 0.1%  | 0.0% | 0.0% | 2006+                     |
| AMNETE  | AmmanNet (English)                              | 474   | 15  | 23  | 3.2%  | 0.0% | 0.0% | 2010-<br>2011,<br>2015+   |
| AFRCA   | Cape Argus (Cape Town)                          | 969   | 9   | 11  | 0.9%  | 0.0% | 0.0% | 2010                      |
| AFRAT   | Arusha Times (Arusha)                           | 2041  | 7   | 10  | 0.3%  | 0.0% | 0.0% | 2010-2016                 |
| BOLEXM  | Bolivian Express Magazine                       | 341   | 1   | 2   | 0.3%  | 0.0% | 0.0% | 2016+                     |
| BAYALY  | Bayane Al Yaoume                                | 4683  | 0   | 0   | 0.0%  | 0.0% | 0.0% | 2016+                     |

|        |                                                     |         |        |        |      |      |      |           |
|--------|-----------------------------------------------------|---------|--------|--------|------|------|------|-----------|
| BIRZAR | Birzha.ru-articles                                  | 1889    | 0      | 0      | 0.0% | 0.0% | 0.0% | 2015-2016 |
| BIRZFW | Birzha-Federal and World News                       | 674     | 0      | 0      | 0.0% | 0.0% | 0.0% | 2015+     |
| BIRZNN | Birzha.ru-news Nizhnii Novgorod and Novgorod region | 2326    | 0      | 0      | 0.0% | 0.0% | 0.0% | 2015+     |
| RGSUBB | Rossiyskaya Gazeta - Weekly                         | 5012    | 0      | 0      | 0.0% | 0.0% | 0.0% | 2015+     |
|        |                                                     | 9370951 | 541081 | 951941 |      |      |      |           |

## Evaluation of the Automated Coding Process

The basic criteria for evaluation of the automated coded process are: quality, time, and cost. That is, automated coding is worthwhile if it produces data equal or better in recall and/or accuracy to manual coding, in the same or shorter time-frame, and at equal or lesser cost.

### *Quality*

Unfortunately, quality assessments of the prior manually coded data do not exist. On the other hand using the development process automated coding schemes that meet or exceed recall  $\geq 0.9$  and accuracy  $\geq 0.6$  can be achieved meeting the desired quality levels. However, we were able to assess the agreement between manually coded and machine coded data for a subset of the variables and the time period: 370 MAROBID\_YEARs for Eastern Europe, 2005/200 and 111 MAROBID\_YEARs for Middle East and North Africa. The low agreement for both samples when a variable is coded in the manual data set is a cause of concern but the differences in the processes, the range of sources, and the documents retrieved for analysis make evaluation difficult. At first glance, the automated coding process appears to be under-coding but that cannot be determined without an out-of-sample assessment of the automated coding process.

### **Correspondence between Automated and Manual Data for 370 MAROBID\_YEARs for Eastern Europe, 2005/2006**

| BARK Variable | MAROB Variable | Agreement Overall | Agreement when <i>not</i> coded in MAROB | Agreement when coded in MAROB |
|---------------|----------------|-------------------|------------------------------------------|-------------------------------|
| LEFT          | LEFTORG        | 87%               | 93%                                      | 20%                           |
| RIGHT         | RIGHTORG       | 88%               | 94%                                      | 20%                           |
| ETHN          | ETHORG         | 51%               | 90%                                      | 7%                            |
| TERRCNTRL     | ORGST9         | 88%               | 89%                                      | 33%                           |
| SYM_RESIS     | ORGST1         | 58%               | 98%                                      | 4%                            |
| VIOL_REPRESS  | STATEVIOLENCE  | 9%                | -                                        | 9%                            |

**Correspondence between Automated and Manual Data for 111 MAROBID\_YEARS for  
Middle East and North Africa**

| BARK Variable | MAROB Variable | Agreement Overall | Agreement when <i>not</i> coded in MAROB | Agreement when coded in MAROB |
|---------------|----------------|-------------------|------------------------------------------|-------------------------------|
| RELI          | RELORG         | 58%               | 58%                                      | 59%                           |
| LEFT          | LEFTORG        | 65%               | 72%                                      | 44%                           |
| RIGHT         | RIGHTORG       | 75%               | 77%                                      | 50%                           |
| NATILST       | NATORG         | 62%               | 90%                                      | 20%                           |
| ETHN          | ETHORG         | 42%               | 63%                                      | 26%                           |
| GENDEXC       | GENDEXC        | 95%               | 100%                                     | 0%                            |
| SUPR          | SUPREMACIST    | 78%               | 100%                                     | -                             |
| ENVI          | ENVORG         | 100%              | 100%                                     | -                             |

*Time*

After the initial investment in coding scheme development the automated document acquisition and coding process is much faster than manual document acquisition and coding with near real-time acquisition and coding possible, reducing lag of 1 to 2 years to as little as 3-4 days. In addition, automated coding allows for increased temporal resolution of the data set from yearly to monthly, weekly, or at the extreme daily, with data coded from millions of documents.

*Cost*

The automated process compares favorably on cost as well as time: the cost for a year of automated document acquisition and coding is \$130,000 per year and is expected to decrease as the systems are refined. In contrast the manual process costs \$160,000 per year and increasing as labor costs increase.

### Cost and scope comparison of automated and manual data collection

|                                                            | <b>Automated</b> | <b>Manual</b> |
|------------------------------------------------------------|------------------|---------------|
| <b>Cost per year of operation (years of data may vary)</b> | \$129,282        | \$155,172     |
| <b># of Organizations collected for</b>                    | 498              | 200           |
| <b># of Observations per Year (AVG)</b>                    | 94,705           | Unknown       |
| <b># of Sources with Observations per Year (AVG)</b>       | 55               | Unknown       |
| <b># of Documents Processed per Year of Data (AVG)</b>     | 1,489,906        | Unknown       |

#### Previous Manual Coding

- ▶ Resource intensive with human coders
  - \$160K/yr and increasing
- ▶ Not anywhere near real time
  - Lag of one or more years
- ▶ Low temporal resolution
  - Yearly

#### New Automated Coding

- ▶ Reduced resources required
  - \$130K/yr and decreasing with additional data management automation.
- ▶ Near real time
  - 7 days of source texts can be coded in 2-3 days.
  - Daily processing possible.
- ▶ Higher temporal resolution
  - Yearly, monthly, weekly, or even daily.
- ▶ High volumes of source texts can be coded
  - Millions of documents.

## ***Hazard Modeling Analysis***

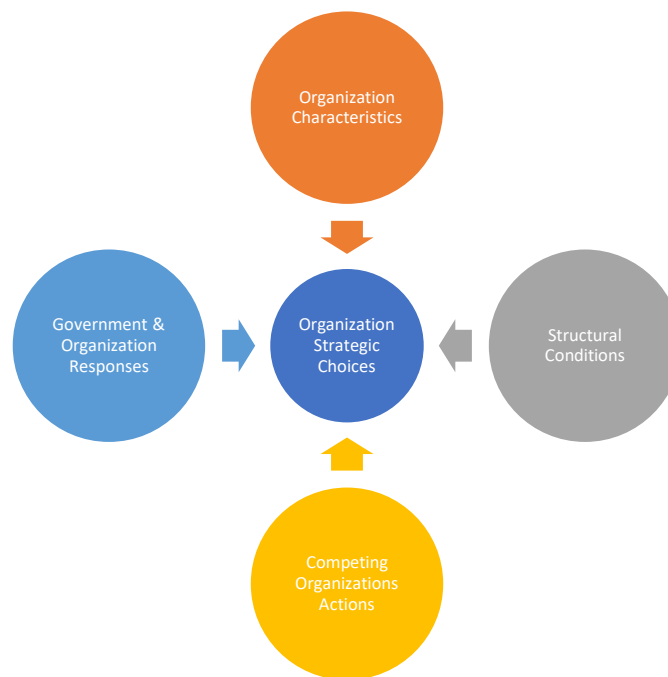
J. Michael Greig  
Department of Political Science  
University of North Texas  
[greig@unt.edu](mailto:greig@unt.edu)

Joseph K. Young  
School of International Service and  
School of Public Affairs  
American University  
[jyoung@american.edu](mailto:jyoung@american.edu)

### Causal Logic: Initiation of Activities

We conceptualize the choice of non-state organizations to initiate protests, violence, and negotiations with their governments as a function of four distinct sets of forces, which we summarize in Figure 1. First, the characteristics of a non-state organization not only condition the goals of the group, but influence the perceived suitability and effectiveness of strategic approaches for the attainment of the goals of the organization. With that in mind, we focus our analysis on the ideological and resource characteristics of the organizations in our sample. We examine the distinct effects that leftist, rightist, religious, and ethnic ideologies play in shaping the likelihood that an organization will undertake violence, take part in negotiations or engage in protests for the first time. Likewise, we explore the effect that control over territory plays in the propensity that an organization will adopt violence, negotiate or participate in protests. Territorial control by an organization both tends to provide a greater pool of resources available to a group for the achievement of its political goals as well as confers a greater sense of legitimacy upon the group among the populace that we see as likely to influence the strategic approaches that such organizations select.

**Figure 1: Influences on Non-State Organization Strategies**



Second, we focus our analysis upon the structural conditions that are present in the environment in which an organization operates. We see these structural conditions as shaping the strategic environment or opportunities in which organizations make choices regarding how best to achieve their political goals. These structural conditions tend to constrain and enable strategic choices among organizations in ways that we see as likely to influence their tendency to undertake violence, engage in protests, or participate in negotiations. Societies in which there are high levels of dissatisfaction with the government and substantial suffering among the populace are

ones in which governments tend to have a lower level of legitimacy among their people and organizations challenging these governments are more likely to see a supportive populace for their challenge. This supportive populace, in turn, tends to create a pool of disaffected people that makes it easier and more fruitful for organizations to undertake violence, negotiate and initiate protests. As such, we include in our models variables describing the quality of life in the societies in which organizations operate and the strength of the state's security apparatus. These forces include measures of per capita GDP, life expectancy, and the size of the country's military force. Likewise, we also look to broader state characteristics that point to higher prospects for challenges in governance that we see as likely to undermine governments ability to govern territory and empower organizations that seek to challenge them. To that end, we include in our models measures describing population size, the level of ethnic fractionalization, and the geographic size of the state in which a non-state organization operates in. Lastly, we also model the effect that the broad level of political permissiveness and opportunity in a state plays in conditioning the behavior of organizations. As such, we examine the effects that the level of democracy and the overall extent of political repression engaged in by the government play in the initiation of protests, violence, and negotiations by organizations.

As our third set of influences on the strategic choices of organizations, we look specifically at how the government responds to the demands of an organization and, in turn, how the prior strategic choices of an organization condition subsequent decisions to initiate protests, violence, and negotiations. With respect to government responses, we focus our attention on the most severe form of government response to the demands of a non-state organization: the use of violent repression against the organization's members. Although the overall level of repression present in a society can condition the strategic choices of non-state organizations, we see an important causal distinction between the overall level of repression in a society and the level of repression directed at a specific group. By way of analogy, this distinction is similar to the economy. For example, the economy could be good but an individual may lose their job. Or the economy could be in recession whereas a person gets promoted. Related to repression, the overall environment may be highly repressive, but a government avoids directly repressing a group (like the Syrian government did for a long time with ISIS). The converse is also true. A government may be quite permissive generally but use harsh repression against a particular group. This distinction highlights the need to examine how both types of repression influence group choice.

In terms of prior group behavior, it makes sense to expect that as non-state actors evaluate their strategic options and make choices regarding the initiation of violence, protests, or negotiations, they are likely to incorporate into their decision-making the prior strategies they have employed. A non-state organization might, for example, pursue an escalatory strategy in which it seeks to steadily increase its pressure on its government by transitioning from demands to protests to negotiations. By contrast, another non-state organization that has successfully brought the government to the negotiating table may see it as less desirable to initiate violence against the government. In this sense, we expect that the prior strategic choices made by non-state organizations will shape their future choices.

Our fourth set of influences centers upon the activities of other non-state organizations operating in a state. A non-state organization does not operate by itself in a vacuum with the government. Instead, most states have a variety of distinct non-state organizations operating within the

country, each pursuing its own goals and seeking to draw attention and extract concessions from the regime. At the same time, this multitude of non-state actors within a country must also compete with one another for members and support. This serves to create an environment in which both the presence and activities of other non-state actors influence the strategic decisions of one another. Non-state actors not only compete with another, but they can also learn from the actions of each other. With this logic in mind, we expect that the number of non-state organizations within a state will shape the competitive environment faced by each organization in ways that influence its strategic choices to initiate protests, violence, and negotiations. At the same time, we also expect that the extent to which other organizations have previously engaged in protests, violence, and negotiations, will influence the extent to which an organization is likely to follow suit.

## ANALYSIS

### *Data*

This analysis focuses upon the initial occurrence of three distinct forms of political engagement between governments and non-state organizations operating within their territory. First, we examine the forces associated with the initiation of protests by non-state organizations. Second, we explore the causal forces linked to the initiation of violence by non-state organizations. Lastly, we analyze the conditions conducive to the initial occurrence of negotiations between a government and a non-state organization. Our decision to explore these three forms of political behavior between non-state organizations and their governments separately is rooted in our expectation that each will carry an important causal link for the other. In this respect, our analysis allows us to explore how the decision to engage in protests shapes the subsequent choice of a non-state organization to undertake violence and, in turn, how the occurrence of protests and violence by non-state organizations conditions the willingness of governments to undertake negotiations with non-state organizations operating within their territory.

The basis for our analysis is data drawn from *The Organization that Does(n't) BARK MENA (BARK\_MAROB) Dataset*, which contains detailed information regarding the characteristics and behavior of ethnopolitical organizations operating in the Middle Eastern and North African region. This data also includes information on the responses of the home governments of the non-state organizations in the data to their activities – including information identifying the use of repression against as well as the initiation of negotiations with a non-state organization. We augment this data with data that provides information on the military capacity of home governments, the form and quality of governance, and the characteristics of the state in which a non-state organization is based. Data on military capacity is drawn from the Correlates of War National Material Capabilities Dataset 5.0 (Singer, Bremer, and Stuckey 1972). Data on governance characteristics was based upon the Quality of Government Standard Dataset (Teorell, Jan et al. 2019).

The unit of analysis used in our data is the non-state organization year. The temporal range for our analysis extends from 2005-2017. Our sample includes a total of 104 non-state organizations across 12 countries. Functionally, as explained below, the number of non-state organizations varies from analysis to analysis depending upon whether or not a non-state organization was

engaged in violence, protests, or negotiations during the year in which it initially enters the dataset.

### *Method*

Because we are interested in the causal forces associated with the *initiation* of each type of political behavior by non-state organizations, we employ hazard analysis as the statistical tool for our analysis.<sup>1</sup> For our analysis, hazard analysis brings two chief virtues. First, hazard analysis is a tool that is specifically built to evaluate the impact of casual variables upon the time until an event – such as the initiation of violence – occurs. Forces that increase the hazard of an event occurring serve to shorten the time until that event takes place. By contrast, forces that reduce the hazard of an event taking place function to increase the duration until an event occurs. Second, hazard analysis also offers the benefit of dealing with censored observations. Censored observations occur when a case persists until the end of the dataset without an event occurring. For example, in our dataset, which terminates in 2017, it is possible that a non-state organization that has remained non-violent throughout the entirety of the time period of our analysis might become violent in some future year, such as 2018, 2022, or 2050. As a result, arbitrarily assigning a date in our analysis for an event that might yet occur but has not yet done so would be incorrect and yield faulty inferences. Fortunately, hazard analysis offers a statistical approach that is specifically designed to correctly address this censored data issue, mitigating the risk of faulty inferences.

Following this approach, we estimate separate hazard models for the initial occurrence of violence, protests, and negotiations by a non-state organization. While hazard analysis offers a wide variety of statistical modeling choices, we employ the most commonly used form of hazard analysis, the Cox regression model. This model imposes the fewest assumptions upon the data and thereby represents a more conservative modeling approach for our data. Given data limitations and the large number of potentially influential variables across our four sets of theorized causal forces, we adopt an iterative approach to building our statistical models. As a result, we first estimate a separate hazard model containing variables for each of the four sets of causal forces specified above. We then use the results from these initial models to build a comprehensive model that includes the strongest predictors of each type of hazard. In the tables below, we report the results from all of these models and discuss the insights gleaned from each comprehensive model.

Because hazard analysis examines the time until an event occurs, the data structure for the analysis requires that actors must enter the data *before* an event occurs. As a result, for each of our analyses, each non-state organization must exist in the data for at least one year before an event – violence, protests, or negotiations – occurs. Because of this requirement, not all non-state organizations in the sample, such as those that are violent at the very beginning of their entry in the dataset, can be included in the analysis. As a result, the number of non-state organizations analyzed varies from analysis to analysis. Our analysis of the initiation of violence contains a sample of 94 non-state organizations, of which 7 become violent at some point in the analysis. Our analysis of protests includes a sample of 71 non-state organizations, of which 21 ultimately

---

<sup>1</sup> Other names for this modeling approach include, survival modeling and event history analysis. The technique originated in biomedical research to examine how different treatments influenced the survival of patients.

initiate protests. Lastly, our analysis of the initiation of negotiations focuses on a sample of 97 non-state organizations, among which 9 ultimately engage in negotiations with their government.

An important assumption of the Cox regression models used in this analysis is that variables exert a consistent effect upon the hazard of an event – protests, violence, or negotiations – over time. As a result, we empirically evaluate each of the models used in our analysis for violations of this assumption. Where we identify violations of this assumption, we apply the standard correction for such violations by interacting the offending variable with time. We report these time effects in the tables below and discuss their implications in the summary of the findings.

## FINDINGS

Our findings highlight that, to an important degree, the initiation of protests, violence, and negotiations tend to be driven by distinct casual processes. In general, the initiation of protests tends to be more directly driven by the organizational characteristics of non-state organizations than the initiation of violence by organizations tends to be. By contrast, the initiation of violence by non-state organizations is better connected to the structural conditions in which a non-state organization operates than the onset of protests or negotiations are. That said, our findings also point to some important commonalities among the drivers of protests, violence, and negotiations by non-state organizations. The way in which a non-state organization's government responds to its grievances plays an important role in shaping the likelihood the organization will engage in protests, negotiate or adopt violence. Likewise, we also find evidence for an important set of cascade effects in which the actions of other non-state organizations within a state influence the likelihood that a non-state organization will undertake protests, violence, or negotiations. Finally, our findings underscore the important extent to which the adoption of protests, violence, and negotiations by non-state organizations impact the occurrence of one another. In this sense, while the initiation of each form of non-state organization behavior is driven by a unique combination of forces, each carries implications for the other.

### *Initiation of Protests*

Protest activity is a common activity among non-state organizations in our data. Among the 104 non-state organizations in the data, 33 undertake protest activity either before or immediately upon their entry into the data. A further 21 non-state organizations that were not previously engaged in protests subsequently undertake protests at some point during the time period under analysis by our model. As Table 1 shows, the organizational characteristics, government response, and activities of other non-state organizations play a significant role in the hazard that a non-state organization initiates protests.

The ideological characteristics of a non-state organization exerts an important impact upon the likelihood that the group will adopt protests. Leftist organizations, for example, are better than four times more likely to initiate protests than other types of non-state organizations. Non-state organizations with a religious ideology are even more likely to engage in protest behavior, showing an eleven-fold higher chance of adopting protests than other types of non-state actors. We, however, see no evidence that non-state organizations with rightist or ethnically-based ideologies are significantly more likely to undertake protests.

The way in which the government in which a non-state organization is based reacts to the demands of the group carries powerful implications for the likelihood that an organization will initiate protests. A non-state organization that is targeted for violent repression by its government is six times more likely to initiate protests than an organization that does not face violent repression. In this sense, repression, rather than dissuading challenges against the government, is more likely to stimulate more aggressive action by non-state organizations with grievances against the regime.

The relationship between a non-state organization and its government does not occur within a vacuum. Instead, our findings highlight that both the presence and actions of other non-state organizations within a country condition the likelihood that a non-state organization will, itself, initiate protests. Each additional non-state organization within a state increases the likelihood that a non-state organization will initiate protests against its government by 23%. This result may suggest that the more competition that a non-state organization faces for attention and resources from its government, the more likely the group is to seek to undertake activities aimed at drawing attention from the regime to its demands. Interestingly, the more that other non-state actors engage in protests within a state the less likely a non-state organization is, itself, to initiate protests. Each additional prior year of protests by another non-state actor in a country reduces the hazard that a non-state organization in that same country will initiate protests by 48%. This result may suggest that, rather than non-state organizations bandwagoning together in protest, organizations may seek other paths to achieve their goals the more that other actors engage in protests.

### *Initiation of Violence*

Violence by non-state actors is, not surprisingly, given both its costs and risks, less commonly engaged in by organizations in our sample of cases. Of the 104 non-state organizations in our data, 10 engage in violence either before or immediately upon entry into the data in our analysis. An additional 7 non-state organizations that were initially non-violent in our data subsequently become violent. The adoption of violence is most directly driven by the environment in which a non-state organization operates, the response of the government to its demands, the activities of other non-state actors, and, importantly, the previous actions of the non-state organization itself.

As Table 2 shows, as was the case with the initiation of protests by non-state organizations, violent repression directed at an organization by its government massively increases the hazard that the group will initiate violence, although this effect abates somewhat over time. In the first year in which a non-state organization is targeted for violent repression by its government, the hazard of it adopting violence increases by a factor of 7688, an enormous increase in the risk of violence. While this increase in hazard from violent repression declines the longer a group experiences this repression without resorting to violence, the risk of the initiation of violence remains extraordinarily high. Even at the 10-year mark of violent repression against the group by its government, the hazard of an organizational resort to violence remains heightened by a factor of 1033. In short, rather than dissuading non-state actors from adopting violence, violent repression by a government against an organization is the single most powerful pressuring force for the initiation of violence in our analysis.

Beyond the use of repression by the government against a non-state organization, the broader structural environment in which a group operates also conditions its likelihood of adopting violence. Non-state organizations operating in countries with a larger number of ethnic divisions are more likely to transition toward the use of violence than those in more ethnically homogenous states. Likewise, non-state organizations in countries with a larger geographic size are also more likely to adopt violence. This effect may reflect both the extent to which larger countries have more frontier areas from which opposition groups can mount campaigns of violence and the greater difficulty a government faces in securing a larger homeland territory. Interestingly, non-state actors in countries with larger populations are less likely to adopt violence.

As was the case in the adoption of protests, the actions of other non-state actors in the home country of an organization also condition its likelihood of initiating violence. Each additional non-state actor in an organization's home state increases its likelihood of adopting violence by 69%. Yet, the more frequently that these other non-state actors engage in protests, the less likely an organization is to initiate violence.

Our findings point to a clear connection between protest by non-state organizations and the adoption of violence. Governments often respond to protests with repression due to a fear of protests spiraling into violence. Yet, our findings not only underscore the degree to which repression heightens the risk of violence, but also highlight that protests can function as an outlet valve against the adoption of violence by non-state organizations. A non-state organization that engages in protests is initially 68% less likely to adopt violence than an organization not involved in protest. The negative effect of protests on the adoption of violence diminishes the longer that a group continues to engage in protests, but protest activity still serves to dampen the hazard that an organization will adopt violence. A non-state organization is still 62% less likely to adopt violence after 5 years of protest and 54% less likely to initiate violence after 10 years of protest than an organization not involved in protests.

Unlike the initiation of protests, we see no evidence that the ideological characteristics of non-state organizations play a role in the adoption of violence. Surprisingly, the military capabilities of the government faced by a non-state actor appears to play no role in the likelihood that an organization will choose to adopt violence. Neither a country's number of military personnel nor its level of military expenditure appears to influence the likelihood that a non-state organization will initiate violence. It is important to recognize, however, that while a government's military capabilities may not dissuade an organization from adopting violence, these capabilities may well have a strong effect upon the ability of a non-state actor to sustain a campaign of violence and accomplish its goals through violence.

### *Initiation of Negotiations*

Negotiations between non-state organizations and governments tend to be relatively rare in our sample of cases. Among the 104 non-state organizations in our data, only 7 experience negotiations either before or when they first enter our data sample. An additional 9 organizations experience negotiations at some point during the period of time upon which our analysis is conducted. The limited amount of negotiations between non-state organizations and governments is not surprising, given the degree to which governments typically seek to avoid granting

legitimacy to domestic opposition groups. The fundamental drivers of negotiations in our analysis are the organizational characteristics of non-state actors, the activities of other non-state actors, and the prior behavior of the non-state organization itself. These findings are summarized in Table 3.

Non-state actors with a rightist or an ethnically-based ideology are significantly more likely to negotiate with their government. The hazard of negotiations initially taking place with the government are nearly 4 times greater for rightist organization and almost 17 times greater for an ethnic organization than organizations with different ideological orientations. We found no evidence that religious or leftist ideologies influenced the likelihood that negotiations would begin with the government. Likewise, the control of territory by a non-state organization, which would seemingly place the group in a more advantageous power position favorable for negotiations, was not significantly linked to the initial occurrence of negotiations with the government.

The more that the government of the home state of non-governmental organization negotiates with other non-state actors, the more likely negotiations are to start with the non-governmental organization itself. Initially, government negotiations with another non-state actor increase the hazard of the start of negotiations with a particular non-state organization by 114%. This increased hazard of negotiations with a particular non-state organization diminishes very slightly, however, the longer negotiations continue with other actors without starting with the non-state organization. Even for a non-state actor that has not seen negotiations with the government after ten years of negotiations between the government and other actors, the hazard of negotiations is still more than double that of group in a state where no negotiations with other actors have occurred. It, however, is unclear on whether it is an overall increased willingness of the government to negotiate with non-state organizations that drives this effect or an improved bargaining position of the non-state organization.

The past activities of a non-state organization also directly shape the likelihood that negotiations will begin with the government. In general, governments show a strong tendency to reward protest activity with talks while punishing violence by avoiding negotiations with violent groups. Each prior year of protest activity by a non-state actor increases the likelihood that the group will see the start of negotiations with the government by nearly 50%. By contrast, each prior year of violence by a non-state organization reduces the likelihood that negotiations will begin with the government by more than 46%. This suggests that it is significantly easier for governments to grant the legitimacy that negotiations confer to groups that remain non-violent but undertake protests than to organizations cross the threshold of normal political activity and engage in violence.

### Forecasting and Machine Learning

While examining the tables, coefficients and effect sizes can be helpful, what we do not know is how predictive these models are from this information. In a perfect world, we would estimate these models on all available data, collect new data and then test these models on the new data. This perfect world is often impractical or impossible. One way to do approximate this impossible task is to *train* the model on a portion of the data, then *test* the model on the piece that was left

out of the estimation. This *out-of-sample* forecasting or what is often termed cross-validation allows us to examine the predictive ability of the model.

There are two tools of machine learning we use to evaluate the ability of the model to predict and properly classify the data. First, we use receiver operator characteristic curves to examine how well the model is classifying ones and zeroes. In this case, that means how well we are, 1. Predicting onsets of protest/no protest, 2. Onsets of violence/no violence and, 3. Onsets of negotiations/no negotiation. Estimating logit models using the same covariates as above, we can then predict the onsets of these different outcomes or occurrences. We do so by generating a score from 0-1 with numbers closer to 1 meaning we are closer to classifying each one and zero accurately. In this type of prediction task 0.7-0.8 is usually seen as acceptable 0.8-0.9 is viewed as good and above 0.9 is seen as excellent. This approach is always viewed in contrast to the base prediction of just always choosing the most likely outcome. In short, since protest happens in only about 31% of the organization years, we can be 69% accurate if we always just guessed 0 or no protest.

Second, we use a machine learning technique called *lasso* (least absolute shrinkage and selection operator). It is a regression technique that helps with variable selection to enhance the predictive accuracy of the model. We use lasso to hone in on the covariates that are the most useful in prediction.

From the cross-validation, we report the predictive accuracy of the model in subsets of the data, MENA and the universe, and in predicting protest and violence and negotiation, and when accounting for time dependency or not (See Table 4). In Table 5, we report the variables from the lasso regression modeling that aid in predictive modeling.

In general, our models are better at predicting onsets or timing of protests than violence. Predicting negotiations falls somewhere in between. This is true in both the MENA data as well as the broader universe data. In the protest data, whether we are predicting onsets of protests or the timing of a protest, on average, the model is excellent. We are predicting onsets and timing of protests with anywhere from 92 to 97% accuracy. Both in the MENA and Universe data, we have high levels of accuracy. There is greater spread in the violence data. In general, our models are better at predicting violence in the MENA data than a more general data set. Our models are also drastically better at producing predictions of timing of violence than onsets. That is, we are better at predicting when violence is likely to occur given that it happened in the past than we are at predicting initial onsets (0.91 as compared to 0.83). Where the model is least effective is predicting violence onsets in the universe data (0.718). Since a group being violent in a year is only about 23% likely, a naïve model would be accurate about 77% of the time. In short, this is the weakest part of the model. Our model is highly accurate at predicting negotiations in the universe data (92 to 93%). But, the model, is less accurate on MENA only data.

In sum, we are confident that this model includes important covariates for explaining and predicting the onset and timing of violence, negotiations and protest. Where it is the least accurate is in predicting onsets of violence in a global sample. This is likely due to regional and country-specific factors.

In the lasso machine learning regression, we identify which variables are most useful for prediction/classification and drop out the least useful ones. This is an alternative way to hone in on critical factors for prediction. Table 5 shows which variables are retained for each model using MENA data: protest, violence, and negotiations. Recall that we grouped the factors in four categories (See Figure 1). Table 3 has each of these characteristics in a column along with a comprehensive model that includes all four (org. characteristics, structural conditions, responses, competing orgs.).

In the protests model, mostly the group characteristics data (ideology, territorial control), the government choice (violent repression directed at the group), and time were useful for prediction. In other words, many of the structural variables were not helpful (outside of population and life expectancy) at predicting protests at the group-year). In the protests in the universe data, the same group characteristics variables were useful. When moving to violence, these structural variables were all helpful in predicting violence as well in both models. In predicting violence, all of the group characteristic variable remained useful as well as the government action variables. The political terror scale, or the more general repressive environment, mattered to explain violent choices by groups whereas it didn't matter for explaining protest. This fits with larger arguments about violence by the government spawning violence by the groups in an action-reaction cycle. Interestingly GDP, which has been a strong predictor of civil war helps predict violence but not protests or negotiation. This suggests that the drivers among protests, negotiation and violence may not be the same or that the underlying motivations for violent and nonviolent mobilization may be quite different. In sum, we were most accurate in predicting protest with fewer organizational level variables accounting for this relative success. With both negotiations and violence, more variables were helpful in prediction but they were less accurate. Across all dependent variables, general repression and directed repression had large effects in helping us predict timing and onsets of these outcomes.

In drawing comparisons among the results of the analysis, it is important to recognize the differences between the organizations included in the global sample of cases and those included in the Middle East and North African data. In our study of the onset of violence and protests, the analyses for both regions focus attention on groups that are initially not engaged in violence or protests, respectively, when they first come under observation in the data. The global sample of cases, however, is drawn from a set of organizations that have previously been engaged in violence. As a result, many of the differences between the results of the analysis of the global sample of cases and those from the Middle East and North Africa is a function of the global sample's larger inclusion of organizations that have a prior history of violent behavior.

Table 1: Hazard Analysis of the Time Until Initiation of Non-State Actor Protests

| VARIABLES                               | Organization<br>Characteristics<br>Model | Structural<br>Conditions<br>Model | Responses<br>Model | Competing<br>Organizations<br>Model | Comprehensive<br>Model |
|-----------------------------------------|------------------------------------------|-----------------------------------|--------------------|-------------------------------------|------------------------|
| Violent repression directed at group    |                                          |                                   | 17.21**<br>(10.57) |                                     | 6.130**<br>(3.709)     |
| Prior violence                          |                                          |                                   | 0.902<br>(0.154)   |                                     |                        |
| Political terror scale                  |                                          | 4.741*<br>(3.238)                 |                    |                                     | 0.673<br>(0.286)       |
| Ethnic Fractionalization                |                                          | 21.21+<br>(35.71)                 |                    |                                     |                        |
| Population                              |                                          | 3.750<br>(5.136)                  |                    |                                     |                        |
| Per capita GDP                          |                                          | 0.484<br>(0.371)                  |                    |                                     |                        |
| Liberal democracy index                 |                                          | 0.0184<br>(0.0696)                |                    |                                     |                        |
| Life expectancy at birth, total (years) |                                          | 1.499<br>(0.441)                  |                    |                                     |                        |
| Land area                               |                                          | 0.507<br>(0.412)                  |                    |                                     |                        |
| Leftist ideology                        | 6.409*<br>(5.309)                        |                                   |                    |                                     | 4.146+<br>(3.101)      |
| Rightist ideology                       | 8.285e+13**<br>(1.099e+14)               |                                   |                    |                                     |                        |
| Religious ideology                      | 5.478*<br>(3.943)                        |                                   |                    |                                     | 11.20**<br>(9.984)     |
| Ethnic ideology                         | 5.427<br>(7.955)                         |                                   |                    |                                     |                        |
| Territorial control                     | 15.03**<br>(8.885)                       |                                   |                    |                                     | 3.506<br>(3.729)       |
| Other group violence                    |                                          |                                   |                    | 1.033<br>(0.0815)                   |                        |
| Other group protests                    |                                          |                                   |                    | 0.654*<br>(0.112)                   | 0.517*<br>(0.143)      |
| Other group negotiations                |                                          |                                   |                    | 1.443<br>(0.523)                    |                        |
| Number of groups                        |                                          |                                   |                    | 1.116*<br>(0.0501)                  | 1.231+<br>(0.162)      |
| Military personnel                      |                                          | 0.273<br>(0.293)                  |                    |                                     |                        |
| Observations                            | 754                                      | 650                               | 754                | 683                                 | 629                    |
| Wald Chi-square                         | 2379                                     | 18.83                             | 21.66              | 7.538                               | 44.57                  |
| Log Psuedolikelihood                    | -50.97                                   | -52.87                            | -61                | -46.45                              | -27.10                 |

|          |    |    |    |    |    |
|----------|----|----|----|----|----|
| Cases    | 71 | 71 | 71 | 71 | 71 |
| Failures | 23 | 19 | 23 | 23 | 21 |

---

Robust standard errors in parentheses

\*\* p<0.01, \* p<0.05, + p<0.1; one-tailed tests where appropriate

Table 2: Hazard Analysis of the Time Until Initiation of Non-State Actor Violence

| VARIABLES                                 | Organization<br>Characteristics<br>Model | Structural<br>Conditions<br>Model | Responses<br>Model | Competing<br>Organizations<br>Model | Comprehensive<br>Model   |
|-------------------------------------------|------------------------------------------|-----------------------------------|--------------------|-------------------------------------|--------------------------|
| Violent repression directed at group      |                                          |                                   | 2.310<br>(2.664)   |                                     | 9,610**<br>(32,952)      |
| Time*Violent repression directed at group |                                          |                                   |                    |                                     | 0.800**<br>(0.0592)      |
| Prior protests                            |                                          |                                   | 1.100<br>(0.0755)  |                                     | 0.311+<br>(0.187)        |
| Time*Prior protests                       |                                          |                                   |                    |                                     | 1.039*<br>(0.0169)       |
| Prior negotiations                        |                                          |                                   | 0.912<br>(0.177)   |                                     |                          |
| Political terror scale                    |                                          | 1.325<br>(1.107)                  |                    |                                     |                          |
| Ethnic Fractionalization                  |                                          | 2.09e-06**<br>(7.81e-06)          |                    |                                     | 1,777*<br>(5,872)        |
| Population                                |                                          | 93.22*<br>(166.9)                 |                    |                                     | 0.00996*<br>(0.0207)     |
| Per capita GDP                            |                                          | 1.644<br>(1.317)                  |                    |                                     |                          |
| Liberal democracy index                   |                                          | 1.35e-06*<br>(8.87e-06)           |                    |                                     | 1.07e-07**<br>(6.52e-07) |
| Life expectancy at birth, total (years)   |                                          | 0.650+<br>(0.153)                 |                    |                                     | 1.421<br>(0.551)         |
| Land area                                 |                                          | 0.0585**<br>(0.0613)              |                    |                                     | 8.692*<br>(8.575)        |
| Military personnel                        |                                          | 1.831<br>(1.813)                  |                    |                                     |                          |
| Leftist ideology                          | 1.210<br>(1.337)                         |                                   |                    |                                     |                          |
| Rightist ideology                         | 1.973<br>(2.294)                         |                                   |                    |                                     |                          |
| Religious ideology                        | 3.542<br>(5.433)                         |                                   |                    |                                     |                          |
| Ethnic ideology                           | 1.041<br>(0.999)                         |                                   |                    |                                     |                          |
| Territorial control                       | 1.924<br>(2.018)                         |                                   |                    |                                     |                          |
| Other group protests                      |                                          |                                   |                    | 0.274**<br>(0.114)                  | 0.371*<br>(0.163)        |
| Other group negotiations                  |                                          |                                   |                    | 2.650*                              | 2.121                    |

|                      |        |        |        |         |         |
|----------------------|--------|--------|--------|---------|---------|
|                      |        |        |        | (1.260) | (3.379) |
| Number of groups     |        |        |        | 1.346** | 1.689*  |
|                      |        |        |        | (0.130) | (0.431) |
| Observations         | 1187   | 1010   | 1,187  | 1,093   | 1006    |
| Wald Chi-square      | 5.474  | 16.65  | 7.823  | 9.854   | 40.62   |
| Log Psuedolikelihood | -14.73 | -11.51 | -16.23 | -13.28  | -9.162  |
| Cases                | 94     | 94     | 94     | 94      | 94      |
| Failures             | 8      | 6      | 8      | 8       | 7       |

Robust standard errors in parentheses

\*\* p<0.01, \* p<0.05, + p<0.1; one-tailed tests where appropriate

Table 3: Hazard Analysis of the Time Until Initiation of Non-State Actor Negotiations

| VARIABLES                               | Organization<br>Characteristics<br>Model | Structural<br>Conditions<br>Model | Responses<br>Model | Competing<br>Organizations<br>Model | Comprehensive<br>Model |
|-----------------------------------------|------------------------------------------|-----------------------------------|--------------------|-------------------------------------|------------------------|
| Violent repression directed at group    |                                          |                                   | 3.663e+17<br>(0)   |                                     |                        |
| Prior violence                          |                                          |                                   | 0.790<br>(0.144)   |                                     | 0.537**<br>(0.103)     |
| Prior protests                          |                                          |                                   | 1.055<br>(0.191)   |                                     | 1.499**<br>(0.212)     |
| Political terror scale                  |                                          | 0.846<br>(0.425)                  |                    |                                     |                        |
| Ethnic Fractionalization                |                                          | 32.19<br>(136.2)                  |                    |                                     |                        |
| Population                              |                                          | 1.086<br>(1.758)                  |                    |                                     |                        |
| Per capita GDP                          |                                          | 0.138*<br>(0.137)                 |                    |                                     | 0.383<br>(0.249)       |
| Liberal democracy index                 |                                          | 752.1<br>(3,992)                  |                    |                                     |                        |
| Life expectancy at birth, total (years) |                                          | 0.921<br>(0.120)                  |                    |                                     |                        |
| Land area                               |                                          | 0.285+<br>(0.212)                 |                    |                                     | 1.166<br>(0.404)       |
| Military personnel                      |                                          | 3.750<br>(4.258)                  |                    |                                     |                        |
| Leftist ideology                        | 0.262<br>(0.215)                         |                                   |                    |                                     | 2.235<br>(2.654)       |
| Rightist ideology                       | 0.128<br>(0.202)                         |                                   |                    |                                     | 3.807*<br>(2.169)      |
| Ethnic ideology                         | 26.32**<br>(29.24)                       |                                   |                    |                                     | 16.76**<br>(14.94)     |
| Territorial control                     | 6.459e+16**<br>(6.209e+16)               |                                   |                    |                                     |                        |
| Other group violence                    |                                          |                                   |                    | 1.216<br>(0.163)                    |                        |
| Other group protests                    |                                          |                                   |                    | 0.769<br>(0.209)                    |                        |
| Other group negotiations                |                                          |                                   |                    | 0.502+<br>(0.204)                   | 2.148+<br>(0.919)      |
| Time*Other group negotiations           |                                          |                                   |                    |                                     | 0.998+<br>(0.000921)   |
| Number of groups                        |                                          |                                   |                    | 1.062<br>(0.0604)                   |                        |

|                      |        |        |        |        |        |
|----------------------|--------|--------|--------|--------|--------|
| Observations         | 1185   | 1008   | 1185   | 1,088  | 1,088  |
| Wald Chi-square      | 16235  | 17.66  | 2.175  | 6.217  | 85.43  |
| Log Psuedolikelihood | -4.055 | -21.16 | -9.122 | -23.12 | -9.407 |
| Cases                | 97     | 97     | 97     | 97     | 97     |
| Failures             | 9      | 9      | 9      | 9      | 9      |

Robust standard errors in parentheses

\*\* p<0.01, \* p<0.05, + p<0.1; one-tailed tests where appropriate

**Table 4: Area under the ROC Curve for subsets of data and different models**

|                         | Area under the Curve |               |             |               |
|-------------------------|----------------------|---------------|-------------|---------------|
|                         | <i>Universe</i>      |               | <i>MENA</i> |               |
|                         | <b>Mean</b>          | <b>95% CI</b> | <b>Mean</b> | <b>95% CI</b> |
| <i>Protests</i>         | 0.931                | 0.92-0.94     | 0.941       | 0.92-0.96     |
| <i>Protests-time</i>    | 0.947                | 0.94-0.95     | 0.944       | 0.94-0.96     |
| <i>Violence</i>         | 0.718                | 0.70-0.73     | 0.832       | 0.80-0.89     |
| <i>Violence-time</i>    | 0.789                | 0.77-0.81     | 0.908       | 0.85-0.96     |
| <i>Negotiations</i>     | 0.925                | 0.91-0.94     | 0.824       | 0.73-0.87     |
| <i>Negotiation-time</i> | 0.932                | 0.92-0.94     | 0.813       | 0.72-0.87     |

**Table 5: Variables Selected from Lasso Machine-Learning Classification**

| <b>Selected Variables from Lasso-MENA</b> |                 |                     |                 |
|-------------------------------------------|-----------------|---------------------|-----------------|
|                                           | <i>Protests</i> | <i>Negotiations</i> | <i>Violence</i> |
| Violent repression directed at group      | X               | X                   | X               |
| Prior violence                            |                 |                     | -               |
| Prior protests                            | -               |                     | X               |
| Political terror scale                    |                 | X                   | X               |
| Ethnic Fractionalization                  |                 | X                   | X               |
| Population                                | X               | X                   | X               |
| Per capita GDP                            |                 |                     | X               |
| Liberal democracy index                   |                 | X                   | X               |
| Life expectancy at birth, total (years)   | X               | X                   | X               |
| Land area                                 |                 | X                   | X               |
| Military personnel                        |                 |                     | X               |
| Leftist ideology                          | X               | X                   | X               |
| Rightist ideology                         | X               | X                   | X               |
| Ethnic ideology                           | X               | X                   |                 |
| Territorial control                       | X               | X                   | X               |
| Other group violence                      |                 | X                   |                 |
| Other group protests                      | X               | X                   | X               |
| Other group negotiations                  |                 | X                   | X               |
| Number of groups                          | X               | X                   | X               |
| Time                                      | X               | X                   | X               |
| Time <sup>2</sup>                         | X               | X                   | X               |
| Time <sup>3</sup>                         |                 | X                   | X               |

## REFERENCES

Singer, J. D, S. Bremer, and J. Stuckey. 1972. “Capability Distribution, Uncertainty, and Major Power War, 1820-1965.” *Peace, war, and numbers*: 19–48.

Teorell, Jan, Stefan Dahlberg, Bo Holmberg, Sören Rothstein, and Richard Alvarado Pachon, Natalia Svensson. 2019. “The Quality of Government Standard Dataset, Version Jan19.”

## ***Causal Pathway Analysis***

Amanda Murdie  
University of Georgia  
[murdie@uga.edu](mailto:murdie@uga.edu)

Causal pathway analysis allows us to assess not only in the direct effect of certain characteristics on non-state organizational violence but the indirect effects of these characteristics on violence through unique intermediary steps. By using a causal pathway approach, we are able to better understand the process through which violence increases, identifying intermediate effects or mediators that can heighten organizational violence.

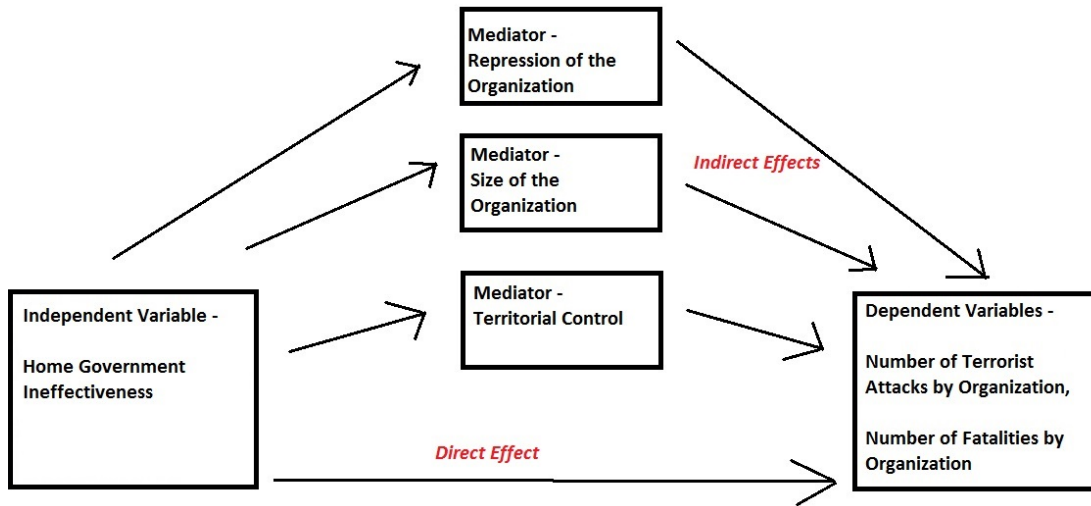
## **Our Causal Model**

Our starting assumption is that effective government will be better able to thwart non-state organizational violence. Government effectiveness is defined consistent with the Worldwide Governance Indicators (WGI) of the World Bank Group. As such, it reflects “perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.” (Kaufmann, Kraay, Mastruzzi 2010, 4). Government ineffectiveness could be thought of as the inverse of government effectiveness: it is the inability of the government to credibly provide the quality services that a population requires. This definition of government ineffectiveness covers a wide variety of public goods that a population could demand, such as health care, education, anti-discrimination policies, or security (Rotberg 2004, 2014). When the government is not providing these services in line with the population’s demands, it will be seen as less capable or effective.

An effective government is one that can accommodate citizen demands credibly. When organizations and individuals make demands to the state, an effective government can consider these demands and formulate a response. An effective government offers responses that are credible; individuals believe the government is going to follow through with its decision. An ineffective government may lack the skills and resources to adjudicate and accommodate demands.

We contend that government ineffectiveness can lead to increased organizational violence both directly and indirectly. Figure 1 outlines this hypothesized process of violence, starting with one independent variable (home government ineffectiveness) and traveling both directly and indirectly to two different dependent variables that capture organizational violence (attacks and fatalities).

**Figure 1: Causal Pathway for Organizational Engagement in Violence**



Directly, as a government becomes more ineffective, organizations may no longer see bargaining with the government as a credible way to garner concessions and address their demands; the only option to increase their concessions and receive their demands is through violence (Fjelde and DeSoysa 2009; Walter 2009). They may also see their likelihood of winning any conflict to be greater when the government is ineffective. The use of violence may greatly alter the organization's bargaining position and lead to the organization usurping power from the government. An increasingly ineffective government may make it difficult to come to credible settlements short of continued violence (Powell 2006; Sobek 2010; Nielsen et al 2011).

Indirectly, we see three potential intermediate steps in the pathway from government ineffectiveness to increased organizational violence: increased government repression directed at the organization, increased organizational size, and increased likelihood that the organization will control territory. First, as a government becomes more ineffective, this will lead the government to use more violent and non-violent repression against the organization, spurring on backlash and increased violence from the organization (Rasler 1996; Walsh and Piazza 2010; Bell et al 2013). Ineffective governments may push their agents to use repression as a form of population control. Accommodation and public goods provision to assuage citizen demands are not likely options, making repression a more viable option when facing increased demands. Further, ineffective governments are less likely to monitor and screen security agents, leading to more repression (Mitchell 2004; Bohara et al 2008). Once this repression ratchets up, organizational violence is likely to follow suit (Walsh and Piazza 2010).

Government ineffectiveness may also indirectly affect organizational violence through its impact on an organization's size. As a government becomes less effective, mobilization to the organization can increase, bringing new individuals with new skills to the organization. As mobilization grows, the organization can screen recruits to carry out more successful and lethal attacks (Bueno de Mesquita 2005). Although organizational size may first increase the skills and resources necessary for violence, there could also be a tipping point. After growing to a certain

size, an organization may have trouble keeping planned violence hidden, limiting the number of successful terrorist events carried out (Berman and Laitin 2008). Additionally, after growing to a certain size, an organization may no longer need violence to increase its bargaining power; it can substitute non-violent means to induce more concessions (Chenoweth and Stephan 2011).

In a similar vein, government ineffectiveness can lead an organization to capture territory, either through a lack of government services to detect territorial capture or through limited incentives for sympathizer audiences to share information on territorial strongholds. Through this territorial control, an organization can better plan attacks, ultimately increasing organizational violence (Fearon and Laitin 2003; Hendrix 2011).

### **Analysis – Universe of Non-State Organizations**

To examine the causal pathways outlined in Figure 1, we first follow the classic approach of Barron and Kenny (1986) and examine:

- (a) the effects of the key independent variable (home government ineffectiveness) on the ultimate outcome of interest (organizational violence, measured by the number of terrorist attacks and number of fatalities),
- (b) the effect of the key independent variable on the mediators (size of the organization, repression, and territorial control), and
- (c) the effect of the mediators on organizational violence when we control for the key independent variable (home government ineffectiveness).

In addition, we make use of recent advances in the study of mediation models by Karlson, Holm, and Breen (2011), Kohler, Karlson, and Holm (2011), and Breen, Karlson, and Holm (2013) and use the Karson, Holm, and Breen (KHB) decomposition method to estimate the extent to which the various mediators explain the relationship between government ineffectiveness and organizational violence.

The key independent variable is the inverse of the government effectiveness measure from the World Bank Group. We use the inverse of government effectiveness to aid in interpretation: as ineffectiveness grows, so should the probability of repression and territorial control, as well as the size of the organization. Additionally, as ineffectiveness grows, so should organizational violence. Government ineffectiveness is recorded for each organization's home government in a given year. In the sample, government ineffectiveness varies from -2.11, indicating an effective government, to 2.06, indicating an ineffective government.

All mediators and dependent variables come from *The Organization that Does(n't) BARK Universe (BARK\_BAAD) Dataset*. The unit of analysis is the organization-year. We include standard control variables on home country characteristics in all models. These controls include the natural log of population and GDP per Capita, from World Development Indicators and UN Statistics respectively. We also include Polity IV's -10 to 10 regime type measure (Marshall, Gurr, and Jagers 2019) and a dichotomous indicator for whether the home country has a conflict

within its borders from the UCDP/PRIO Armed Conflict Dataset Version 19.1 (Gleditsch et al 2002; Pettersson, Hogbladh, and Oberg 2019). All independent variables are lagged one year.

We include two dichotomous indicators for organizations from two countries in the dataset: Turkey and Palestine. Palestine is an interesting case; base country information comes from Israel instead of Palestine. However, it is reasonable to expect that Israeli government conditions do not apply to Palestine. Turkey is included as a dichotomous indicator because of its subsequent outlier status when focusing on Middle East and North Africa (MENA) organizations. Results for the universe dataset are robust to the exclusion of these dichotomous indicators.

Table 1 provides the results of the first stage of the causal process: whether the key independent variable of interest, government ineffectiveness, affects organizational violence. For these specifications, we run generalized estimating equation (GEE) negative binomial models with AR(1) correlation structures and robust standard errors. As shown, for both the dependent variable of the count of terrorist incidents and the dependent variable of the count of fatalities, an organization's violence in a given year increases as home governments become more ineffective.

**Table 1: How Government Ineffectiveness Influences Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                  | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities |
|----------------------------|-------------------------------|--------------------------------|
| Government Ineffectiveness | 0.805*<br>(0.454)             | 1.958***<br>(0.472)            |
| Population (ln)            | 0.0925<br>(0.0958)            | 0.238**<br>(0.114)             |
| GDP per Capita (ln)        | 0.115<br>(0.331)              | 0.422<br>(0.334)               |
| Turkey                     | 1.170<br>(0.791)              | 1.640**<br>(0.735)             |
| Palestine                  | 0.769<br>(0.499)              | 1.543**<br>(0.681)             |
| Regime Type                | 0.00571<br>(0.0166)           | -0.0389**<br>(0.0188)          |
| Conflict Location          | -0.0231<br>(0.145)            | 0.515***<br>(0.133)            |
| Constant                   | -1.306<br>(3.976)             | -6.314<br>(4.151)              |
| Observations               | 5,876                         | 5,876                          |
| Number of Organizations    | 516                           | 516                            |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 2 provides the results of the next stage in the causal process: how government ineffectiveness affects my proposed mediators. As shown, three of the proposed mediators, violent repression, non-violent repression, and territorial control, are influenced by government ineffectiveness in the hypothesized way. The model concerning the other proposed mediator (organizational size) does not reach standard levels of statistical significance.

**Table 2: How Government Ineffectiveness Influences Repression, Territorial Control, and Organizational Size, GEE Models with AR(1) Correlation Process**

| VARIABLES                  | (1)<br>Violent<br>Repression<br>(Logit) | (2)<br>Non-<br>Violent<br>Repression<br>(Logit) | (3)<br>Territorial<br>Control<br>(Logit) | (4)<br>Organizat<br>ional Size |
|----------------------------|-----------------------------------------|-------------------------------------------------|------------------------------------------|--------------------------------|
| Government Ineffectiveness | 0.364***<br>(0.138)                     | 0.234*<br>(0.139)                               | 0.427***<br>(0.136)                      | -0.267<br>(0.612)              |
| Population (ln)            | 0.0450<br>(0.0495)                      | 0.0892*<br>(0.0527)                             | 0.0345<br>(0.0518)                       | -0.212<br>(0.273)              |
| GDP per Capita (ln)        | 0.114<br>(0.0863)                       | 0.0628<br>(0.0857)                              | 0.0736<br>(0.0848)                       | -0.681<br>(0.547)              |
| Turkey                     | -0.0735<br>(0.560)                      | 0.0848<br>(0.508)                               | -0.310<br>(0.642)                        | -0.156<br>(0.944)              |
| Palestine                  | 0.349<br>(0.449)                        | 0.152<br>(0.451)                                | 0.319<br>(0.473)                         | -0.393<br>(0.690)              |
| Regime Type                | 0.0149<br>(0.0104)                      | 0.0212*<br>(0.0116)                             | 0.0181*<br>(0.00995)                     | 0.0432<br>(0.125)              |
| Conflict Location          | 0.222***<br>(0.0764)                    | 0.127*<br>(0.0747)                              | 0.214**<br>(0.0843)                      | 0.271<br>(1.048)               |
| Constant                   | -2.831**<br>(1.314)                     | -3.347**<br>(1.359)                             | -2.473*<br>(1.303)                       | 10.66<br>(8.769)               |
| Observations               | 6,392                                   | 6,392                                           | 6,392                                    | 6,392                          |
| Number of Organizations    | 516                                     | 516                                             | 516                                      | 516                            |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 3 provides the final stage of the Barron and Kenny (1986) framework for investigating causal processes: whether the mediators affect the outcome variables once government ineffectiveness is controlled for. First, worth noting, when the mediators are included in the same model as government ineffectiveness, we still see the positive and statistically significant relationship we saw in Table 1 in columns 2 and 4; government ineffectiveness is in the expected direction but does not reach conventional levels of statistical significance in columns 1 and 3. Further, as shown in columns 1 and 2 in this table, the violent

repression and non-violent repression variables do show the expected relationship. Territorial control is statistically significant in the model in column 2. In both columns 1 and 2, organizational size has an unexpected negative relationship with violence. Further models were run including a squared organizational size variable, reflecting a likely inverse-U relationship with violence: organizational size is associated with increased violence up to a certain tipping point. After that point, organizational size negatively influences violence. When this squared term is included in the model, as shown in columns 3 and 4, we see the expected inverse-U relationship.

**Table 3: How Government Ineffectiveness and the Various Mediators Influence Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                   | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities |
|-----------------------------|-------------------------------|--------------------------------|-------------------------------|--------------------------------|
| Government Ineffectiveness  | 0.701<br>(0.446)              | 2.044***<br>(0.515)            | 0.699<br>(0.449)              | 2.048***<br>(0.517)            |
| Violent Repression          | 0.253**<br>(0.118)            | 1.465***<br>(0.324)            | 0.252**<br>(0.118)            | 1.423***<br>(0.316)            |
| Non-Violent Repression      | 0.313***<br>(0.0917)          | 0.487*<br>(0.274)              | 0.316***<br>(0.0929)          | 0.478*<br>(0.273)              |
| Territorial Control         | 0.180<br>(0.160)              | 0.843***<br>(0.220)            | 0.188<br>(0.159)              | 0.829***<br>(0.217)            |
| Organizational Size         | -0.000173**<br>(7.42e-05)     | -0.00178***<br>(0.000463)      | 0.00526**<br>(0.00265)        | 0.0220***<br>(0.00690)         |
| Organizational Size Squared |                               |                                | -5.55e-06**<br>(2.70e-06)     | -2.29e-05***<br>(7.45e-06)     |
| Population (ln)             | 0.0884<br>(0.0924)            | 0.379***<br>(0.126)            | 0.0962<br>(0.0928)            | 0.379***<br>(0.127)            |
| GDP per Capita (ln)         | 0.0979<br>(0.321)             | 0.517**<br>(0.202)             | 0.103<br>(0.322)              | 0.520**<br>(0.202)             |
| Turkey                      | 1.002<br>(0.766)              | 0.921<br>(0.588)               | 0.985<br>(0.768)              | 0.916<br>(0.588)               |
| Palestine                   | 0.897*<br>(0.506)             | 3.396***<br>(0.985)            | 0.911*<br>(0.507)             | 3.402***<br>(0.988)            |
| Regime Type                 | -0.000745<br>(0.0198)         | -0.0960***<br>(0.0240)         | -0.00183<br>(0.0200)          | -0.0965***<br>(0.0240)         |
| Conflict Location           | -0.0160<br>(0.158)            | 0.817<br>(0.523)               | -0.00844<br>(0.159)           | 0.816<br>(0.524)               |
| Constant                    | -1.574<br>(3.810)             | -11.32***<br>(3.135)           | -1.778<br>(3.823)             | -11.36***<br>(3.150)           |
| Observations                | 5,876                         | 5,876                          | 5,876                         | 5,876                          |
| Number of Organizations     | 516                           | 516                            | 516                           | 516                            |

Tables 4 and 5 provide the results of the KHB method. We first focus on each mediator independently and then include models with all mediators together. All control variables are included in all models but not shown in the summary tables. As shown, government ineffectiveness has both a direct and an indirect effect on organizational violence. Indirectly, we see that government ineffectiveness works through the mediators of violent repression, non-violent repression, and territorial control. The results for the mediator of organizational size do not reach the standard level of statistical significance. Overall, however, as shown in the last two columns of Tables 4 and 5, all the mediators do transmit an indirect effect of government ineffectiveness on organizational violence, as hypothesized and illustrated in Figure 1.

The KHB method allows us to first examine the mediators as a collective group (as shown in the last two columns of Tables 4 and 5 and then disentangle the various contributions of each mediators. When we do this, we see that the indirect effects of government ineffectiveness on organizational violence are channeled mainly through territorial control (42% of the indirect effects), violent repression (35.5%), and non-violent repression (22.5%). Organizational size makes a negligible contribution to the indirect effect.

In short, through using causal mediation models, we can better understand the pathway through which organizations increase their violence. An ineffective government can both directly influence violence and can indirectly influence violence through how ineffectiveness increases the likelihood that a group has control of territory and the likelihood that the government uses violent and non-violent repression. Focusing on improving credible public goods provision and bureaucratic quality can reduce organizational violence.

**Table 4: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Incidents via the Various Mediators, KHB Method with Robust Standard Errors**

|                 | Violent<br>Repression | Non-<br>Violent<br>Repression | Territorial<br>Control | Organizational<br>Size | Organizational<br>Size & Size<br>Squared | All Mediators       | All Mediators<br>(including<br>Organizational<br>Size<br>Squared) |
|-----------------|-----------------------|-------------------------------|------------------------|------------------------|------------------------------------------|---------------------|-------------------------------------------------------------------|
| Total Effect    | 7.801<br>(2.900)**    | 7.801<br>(2.896)**            | 7.801<br>(2.900)**     | 7.801<br>(2.919)**     | 7.801<br>(2.919)**                       | 7.801<br>(2.893)**  | 7.801<br>(2.893)**                                                |
| Direct Effect   | 6.649<br>(2.804)*     | 7.106<br>(2.825)*             | 6.364<br>(2.776)*      | 7.803<br>(2.919)**     | 7.791<br>(2.918)**                       | 6.442<br>(2.787)*   | 6.441<br>(2.787)*                                                 |
| Indirect Effect | 1.152<br>(0.259)***   | 0.695<br>(0.229)**            | 1.437<br>(0.286)***    | -0.002<br>(0.007)      | 0.010<br>(0.025)                         | 1.358<br>(0.292)*** | 1.360<br>(0.292)***                                               |
| <i>N</i>        | 5,876                 | 5,876                         | 5,876                  | 5,876                  | 5,876                                    | 5,876               | 5,876                                                             |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**Table 5: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Fatalities via the Various Mediators, KHB Method with Robust Standard Errors**

|                 | Violent<br>Repression | Non-Violent<br>Repression | Territorial<br>Control | Organization<br>al Size | Organizationa<br>l Size & Size<br>Squared | All Mediators        | All Mediators<br>(including<br>Organizational<br>Size Squared) |
|-----------------|-----------------------|---------------------------|------------------------|-------------------------|-------------------------------------------|----------------------|----------------------------------------------------------------|
| Total Effect    | 57.786<br>(20.656)**  | 57.786<br>(20.631)**      | 57.786<br>(20.655)**   | 57.786<br>(20.733)**    | 57.786<br>(20.734)**                      | 57.786<br>(20.625)** | 57.786<br>(20.627)**                                           |
| Direct Effect   | 53.127<br>(19.878)**  | 54.725<br>(20.078)**      | 51.834<br>(19.649)**   | 57.790<br>(20.734)**    | 57.754<br>(20.731)**                      | 52.388<br>(19.754)** | 52.372<br>(19.751)**                                           |
| Indirect Effect | 4.659<br>(1.245)***   | 3.061<br>(1.100)**        | 5.952<br>(1.464)***    | -0.004<br>(0.017)       | 0.032<br>(0.074)                          | 5.398<br>(1.431)***  | 5.413<br>(1.435)***                                            |
| <i>N</i>        | 5,876                 | 5,876                     | 5,876                  | 5,876                   | 5,876                                     | 5,876                | 5,876                                                          |

### Analysis – Eastern Europe

With these results in mind, we now turn to replicating the models for only those organizations that have their home base in the Eastern European region (BARK\_MAROB EE). There are a few things to note with this smaller dataset. First, the incidents of terrorism are quite rare; 99.63% of the sample are zeros; there are only 11 1's and 2 2's. The fatalities dataset is also heavily skewed towards zero. Further, incidents only occurred for organizations with a home base in Russia or Uzbekistan. We dichotomize the violence variables and focus on logit models; we also run robustness checks with rare-event logits. Given this heavy skew however, these models should be interpreted with caution.

Tables 1-EE through 5-EE replicate the above analyses but with only the Eastern European sample. As can be seen, the general result as to the direct effect of government ineffectiveness and the relationship between incidents and the mediators continues to hold in many instances. The relationship between government ineffectiveness and the mediators, however, does not hold. As mentioned, these results should be seen as provisional given the very small number of incidents in the dataset, together with the very small number of countries with incidents. Worth noting, results when territorial control was included had trouble converging, especially with the fatalities dependent variable.

**Table 1 – EE: How Government Ineffectiveness Influences Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                  | (1)<br>Terrorist Incidents<br>(Binary) | (2)<br>Terrorist Fatalities<br>(Binary) |
|----------------------------|----------------------------------------|-----------------------------------------|
| Government Ineffectiveness | 8.657***<br>(2.699)                    | 5.924<br>(5.007)                        |
| Population (ln)            | 23.12<br>(30.08)                       | 14.31<br>(25.76)                        |
| GDP per Capita (ln)        | -0.968<br>(5.506)                      | 0.280<br>(1.295)                        |
| Regime Type                | -0.977<br>(0.678)                      | -0.693<br>(0.714)                       |
| Conflict Location          | -22.29<br>(27.60)                      | -14.49<br>(27.53)                       |
| Constant                   | -407.5<br>(485.3)                      | -261.8<br>(444.0)                       |
| Observations               | 2,256                                  | 2,256                                   |
| Number of Organizations    | 188                                    | 188                                     |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 2 - EE: How Government Ineffectiveness Influences Repression, Territorial Control, and Organizational Size, GEE Models with AR(1) Correlation Process**

| VARIABLES                  | (1)<br>Violent<br>Repression | (2)<br>Non-<br>Violent<br>Repression | (3)<br>Territorial<br>Control | (4)<br>Organizational<br>Size |
|----------------------------|------------------------------|--------------------------------------|-------------------------------|-------------------------------|
| Government Ineffectiveness | -0.854**<br>(0.425)          | -0.244<br>(0.525)                    | -0.345<br>(0.330)             | -0.925<br>(0.799)             |
| Population (ln)            | 0.395*<br>(0.231)            | 0.213<br>(0.247)                     | 0.114<br>(0.195)              | -0.172<br>(0.248)             |
| GDP per Capita (ln)        | -0.961***<br>(0.371)         | -0.980**<br>(0.417)                  | -0.779**<br>(0.306)           | -0.779<br>(0.623)             |
| Regime Type                | -0.0407<br>(0.0419)          | -0.0356<br>(0.0404)                  | -0.0196<br>(0.0401)           | -0.0232<br>(0.0200)           |
| Conflict Location          | -0.616**<br>(0.254)          | -0.0905<br>(0.0871)                  | -0.0141<br>(0.0796)           | 1.460<br>(1.575)              |
| Constant                   | -0.209<br>(3.689)            | 2.514<br>(3.621)                     | 2.604<br>(3.319)              | 9.766<br>(8.961)              |
| Observations               | 2,444                        | 2,444                                | 2,444                         | 2,444                         |
| Number of Organizations    | 188                          | 188                                  | 188                           | 188                           |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 3 – EE : How Government Ineffectiveness and the Various Mediators Influence Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                   | (1)<br>Terrorist<br>Incidents<br>(Binary) | (2)<br>Terrorist<br>Fatalities<br>(Binary) | (3)<br>Terrorist<br>Incidents<br>(Binary) | (4)<br>Terrorist Fatalities<br>(Binary) |
|-----------------------------|-------------------------------------------|--------------------------------------------|-------------------------------------------|-----------------------------------------|
| Government Ineffectiveness  | 10.49***<br>(2.199)                       | 8.307*<br>(4.729)                          | 10.06***<br>(2.550)                       | 8.433*<br>(5.058)                       |
| Violent Repression          | 2.930***<br>(0.765)                       | 0.943<br>(1.253)                           | 2.755***<br>(0.465)                       | 0.901<br>(1.244)                        |
| Non-Violent Repression      | -0.0302<br>(2.399)                        | -2.250***<br>(0.705)                       | -0.164<br>(2.228)                         | -2.282***<br>(0.659)                    |
| Territorial Control         | 0.596<br>(1.199)                          |                                            | 0.642<br>(0.844)                          |                                         |
| Organizational Size         | 0.936***<br>(0.261)                       | 0.814<br>(0.581)                           | 2.973***<br>(1.081)                       | 2.260*<br>(1.373)                       |
| Population (ln)             | 78.88***<br>(21.99)                       | 80.18**<br>(33.30)                         | 70.98***<br>(23.86)                       | 76.62**<br>(33.04)                      |
| GDP per Capita (ln)         | -16.01***<br>(4.444)                      | -15.91***<br>(4.866)                       | -14.26***<br>(5.142)                      | -14.88***<br>(4.527)                    |
| Regime Type                 | -1.833***<br>(0.463)                      | -1.823**<br>(0.804)                        | -1.666***<br>(0.504)                      | -1.756**<br>(0.823)                     |
| Conflict Location           | -67.41***<br>(19.79)                      | -69.40**<br>(31.75)                        | -60.69***<br>(21.10)                      | -66.69**<br>(31.80)                     |
| Organizational Size Squared |                                           |                                            | -0.349<br>(0.276)                         | -0.241<br>(0.199)                       |
| Constant                    | -1,270***<br>(352.8)                      | -1,288**<br>(550.1)                        | -1,148***<br>(380.4)                      | -1,236**<br>(549.5)                     |
| Observations                | 2,256                                     | 191                                        | 2,256                                     | 191                                     |
| Number of Organizations     | 188                                       | 29                                         | 188                                       | 29                                      |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 4 - EE: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Incidents via the Various Mediators, KHB Method with Robust Standard Errors**

|                 | Violent<br>Repression | Non-Violent<br>Repression | Territorial<br>Control | Organizational<br>Size | Organizational Size<br>& Size<br>Squared | All<br>Mediators   | All Mediators<br>(including<br>Organizational<br>Size Squared) |
|-----------------|-----------------------|---------------------------|------------------------|------------------------|------------------------------------------|--------------------|----------------------------------------------------------------|
| Total Effect    | 11.128<br>(5.924)     | 9.942<br>(4.349)*         | 9.066<br>(3.994)*      | 6.651<br>(3.754)       | 161.253<br>(59.321)**                    | 9.077<br>(4.076)*  | 91.865<br>(77.308)                                             |
| Direct Effect   | 11.388<br>(5.904)     | 10.311<br>(4.376)*        | 9.234<br>(3.991)*      | 7.993<br>(3.868)*      | 7.988<br>(3.570)*                        | 10.307<br>(4.252)* | 10.081<br>(4.581)*                                             |
| Indirect Effect | -0.260<br>(0.128)*    | -0.369<br>(0.149)*        | -0.168<br>(0.100)      | -1.342<br>(1.108)      | 153.265<br>(164.040)                     | -1.230<br>(0.940)  | 81.784<br>(112.492)                                            |
| <i>N</i>        | 2,256                 | 2,256                     | 2,256                  | 2,256                  | 2,256                                    | 2,256              | 2,256                                                          |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**Table 5 - EE: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Fatalities via the Various Mediators, KHB Method with Robust Standard Errors**

|          | Violent Repression | Organizational<br>Size | Organizational Size<br>& Size<br>Squared | All<br>Mediators     | All Mediators<br>(including<br>Organizational<br>Size Squared) | Organizational<br>Size |
|----------|--------------------|------------------------|------------------------------------------|----------------------|----------------------------------------------------------------|------------------------|
| Reduced  | 7.638<br>(5.183)   | 7.325<br>(5.037)       | 5.513<br>(3.390)                         | 108.375<br>(65.051)  | 6.993<br>(3.786)                                               | 89.974<br>(62.083)     |
| Full     | 7.855<br>(5.123)   | 7.645<br>(5.039)       | 6.810<br>(3.465)*                        | 6.322<br>(3.471)     | 8.246<br>(4.030)*                                              | 6.849<br>(4.057)       |
| Diff     | -0.217<br>(0.131)  | -0.320<br>(0.146)*     | -1.298<br>(1.083)                        | 102.053<br>(121.388) | -1.253<br>(1.076)                                              | 83.125<br>(104.824)    |
| <i>N</i> | 2,256              | 2,256                  | 2,256                                    | 2,256                | 2,256                                                          | 2,256                  |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

## Analysis – Middle East and North Africa

The models were also run using the dataset that restricts the sample to only organizations in the Middle East and North Africa (BARK\_MAROB\_MENA). A few caveats are in order concerning this dataset. First, organizations like Hamas are repeated in the dataset with the same home base listed as the Palestinian Territories and then the same dependent variable values; more examination could be necessary. Second, Turkish and Palestinian organizations are potential outliers; dichotomous indicators are used to account for their unique dynamics. And, finally, this is a much smaller dataset. Because of this, these results should also be interpreted with caution.

The Tables 1-MENA through Tables 5 – MENA are provided below. Although the results concerning government ineffectiveness typically fail to meet conventional levels of statistical significance, we continue to find that the hypothesized mediators do affect organizational violence.

**Table 1 – MENA: How Government Ineffectiveness Influences Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                  | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities |
|----------------------------|-------------------------------|--------------------------------|
| Government Ineffectiveness | 0.504<br>(0.733)              | 0.734<br>(1.088)               |
| Population (ln)            | -0.765**<br>(0.382)           | -0.542<br>(0.439)              |
| GDP per Capita (ln)        | -1.388<br>(1.078)             | -1.306<br>(1.306)              |
| Turkey                     | 8.581***<br>(0.945)           | 7.135***<br>(1.327)            |
| Palestine                  | 6.960***<br>(0.938)           | 5.989***<br>(0.767)            |
| Regime Type                | 0.0235<br>(0.0445)            | 0.0578<br>(0.0678)             |
| Conflict Location          | -0.973***<br>(0.178)          | -0.309<br>(0.253)              |
| Constant                   | 22.53<br>(14.63)              | 18.96<br>(15.84)               |
| Observations               | 1,192                         | 1,192                          |
| Number of Organizations    | 111                           | 111                            |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 2 - MENA: How Government Ineffectiveness Influences Repression, Territorial Control, and Organizational Size, GEE Models with AR(1) Correlation Process**

| VARIABLES                  | (1)<br>Violent<br>Repression | (2)<br>Non-<br>Violent<br>Repression | (3)<br>Territorial<br>Control | (4)<br>Organizational<br>Size |
|----------------------------|------------------------------|--------------------------------------|-------------------------------|-------------------------------|
| Government Ineffectiveness | -0.339<br>(0.401)            | -0.663<br>(0.404)                    | -0.728*<br>(0.441)            | -0.398<br>(0.246)             |
| Population (ln)            | -0.252<br>(0.211)            | -0.355<br>(0.228)                    | -0.207<br>(0.219)             | -0.122<br>(0.0951)            |
| GDP per Capita (ln)        | -0.564<br>(0.455)            | -0.945**<br>(0.458)                  | -0.960**<br>(0.463)           | -0.579**<br>(0.264)           |
| Turkey                     | 1.566<br>(1.081)             | 2.347**<br>(1.167)                   | 1.222<br>(1.062)              | 1.048<br>(0.757)              |
| Palestine                  | 0.503<br>(0.777)             | 0.749<br>(0.852)                     | 0.173<br>(0.754)              | 0.0947<br>(0.340)             |
| Regime Type                | 0.00782<br>(0.0292)          | -0.00169<br>(0.0301)                 | 0.0240<br>(0.0320)            | 0.00972<br>(0.0138)           |
| Conflict Location          | 0.213<br>(0.168)             | 0.142<br>(0.199)                     | 0.315<br>(0.216)              | 0.00442<br>(0.0969)           |
| Constant                   | 7.753<br>(6.111)             | 12.55*<br>(6.659)                    | 10.42*<br>(6.145)             | 7.864***<br>(2.949)           |
| Observations               | 1,303                        | 1,303                                | 1,303                         | 1,303                         |
| Number of Organizations    | 111                          | 111                                  | 111                           | 111                           |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 3 – MENA : How Government Ineffectiveness and the Various Mediators Influence Organizational Violence, GEE Negative Binomial Model with AR(1) Correlation Structure**

| VARIABLES                   | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities | (1)<br>Terrorist<br>Incidents | (2)<br>Terrorist<br>Fatalities |
|-----------------------------|-------------------------------|--------------------------------|-------------------------------|--------------------------------|
| Government Ineffectiveness  | 0.909<br>(0.756)              | 1.560<br>(1.190)               | 0.954<br>(0.675)              | 1.674<br>(1.106)               |
| Violent Repression          | 0.530**<br>(0.226)            | -0.747<br>(0.456)              | 0.168<br>(0.242)              | -1.035**<br>(0.507)            |
| Non-Violent Repression      | 0.729***<br>(0.272)           | 2.029***<br>(0.330)            | 0.681**<br>(0.288)            | 1.791***<br>(0.315)            |
| Territorial Control         | 0.944***<br>(0.286)           | 0.789**<br>(0.375)             | 0.648**<br>(0.260)            | 0.384<br>(0.380)               |
| Organizational Size         | 0.148**<br>(0.0688)           | 0.386***<br>(0.0785)           | 1.105***<br>(0.238)           | 1.765***<br>(0.243)            |
| Organizational Size Squared |                               |                                | -0.227***<br>(0.0564)         | -0.332***<br>(0.0647)          |
| Population (ln)             | -0.638*<br>(0.384)            | 0.0688<br>(0.354)              | -0.554<br>(0.383)             | 0.124<br>(0.350)               |
| GDP per Capita (ln)         | -0.764<br>(1.079)             | -0.224<br>(1.401)              | -0.699<br>(0.981)             | -0.157<br>(1.309)              |
| Turkey                      | 6.842***<br>(1.079)           | 3.985***<br>(1.323)            | 6.743***<br>(1.094)           | 4.064***<br>(1.286)            |
| Palestine                   | 7.528***<br>(0.921)           | 6.811***<br>(0.789)            | 7.584***<br>(0.869)           | 7.068***<br>(0.765)            |
| Regime Type                 | 0.00441<br>(0.0499)           | 0.0565<br>(0.0833)             | 0.00364<br>(0.0453)           | 0.0536<br>(0.0814)             |
| Conflict Location           | -0.871***<br>(0.178)          | -0.453***<br>(0.130)           | -1.002***<br>(0.145)          | -0.619***<br>(0.195)           |
| Constant                    | 13.60<br>(14.51)              | -2.056<br>(14.78)              | 11.71<br>(13.64)              | -3.601<br>(13.68)              |
| Observations                | 1,192                         | 1,192                          | 1,192                         | 1,192                          |
| Number of Observations      | 111                           | 111                            | 111                           | 111                            |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 4 – MENA: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Incidents via the Various Mediators, KHB Method with Robust Standard Errors**

|                 | Violent<br>Repression | Non-Violent<br>Repression | Territorial<br>Control | Organizational<br>Size | Organizational<br>Size & Size<br>Squared | All Mediators      | All Mediators<br>(including<br>Organizational<br>Size Squared) |
|-----------------|-----------------------|---------------------------|------------------------|------------------------|------------------------------------------|--------------------|----------------------------------------------------------------|
| Total Effect    | 1.106<br>(0.729)      | 1.106<br>(0.729)          | 1.106<br>(0.730)       | 1.106<br>(0.733)       | 1.106<br>(0.731)                         | 1.106<br>(0.731)   | 1.106<br>(0.731)                                               |
| Direct Effect   | 1.737<br>(0.918)      | 1.787<br>(0.922)          | 1.935<br>(0.987)*      | 1.762<br>(0.941)       | 1.963<br>(0.995)*                        | 2.067<br>(1.027)*  | 2.091<br>(1.034)*                                              |
| Indirect Effect | -0.632<br>(0.271)*    | -0.681<br>(0.287)*        | -0.829<br>(0.350)*     | -0.656<br>(0.301)*     | -0.857<br>(0.350)*                       | -0.961<br>(0.398)* | -0.986<br>(0.402)*                                             |
| <i>N</i>        | 1,192                 | 1,192                     | 1,192                  | 1,192                  | 1,192                                    | 1,192              | 1,192                                                          |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**Table 5 - MENA: Total, Direct, and Indirect Effects of Government Ineffectiveness on Terrorist Fatalities via the Various Mediators, KHB Method with Robust Standard Errors**

|                 | Violent<br>Repression | Non-Violent<br>Repression | Territorial<br>Control | Organizational<br>Size | Organizational<br>Size & Size<br>Squared | All<br>Mediators   | All Mediators<br>(including<br>Organizational<br>Size<br>Squared) |
|-----------------|-----------------------|---------------------------|------------------------|------------------------|------------------------------------------|--------------------|-------------------------------------------------------------------|
| Total Effect    | 1.993<br>(1.220)      | 1.993<br>(1.220)          | 1.993<br>(1.222)       | 1.993<br>(1.227)       | 1.993<br>(1.222)                         | 1.993<br>(1.224)   | 1.993<br>(1.223)                                                  |
| Direct Effect   | 2.914<br>(1.546)      | 3.006<br>(1.555)          | 3.254<br>(1.669)       | 3.002<br>(1.562)       | 3.361<br>(1.689)*                        | 3.457<br>(1.727)*  | 3.546<br>(1.759)*                                                 |
| Indirect Effect | -0.921<br>(0.445)*    | -1.014<br>(0.473)*        | -1.262<br>(0.586)*     | -1.010<br>(0.476)*     | -1.368<br>(0.605)*                       | -1.464<br>(0.656)* | -1.554<br>(0.687)*                                                |
| <i>N</i>        | 1,192                 | 1,192                     | 1,192                  | 1,192                  | 1,192                                    | 1,192              | 1,192                                                             |

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

## References

- Baron, Reuben M., and David A. Kenny. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of personality and social psychology* 51, no. 6 (1986): 1173.
- Bell, Sam R., David Cingranelli, Amanda Murdie, and Alper Caglayan. "Coercion, capacity, and coordination: Predictors of political violence." *Conflict Management and Peace Science* 30, no. 3 (2013): 240-262.
- Berman, Eli, and David D. Laitin. "Religion, terrorism and public goods: Testing the club model." *Journal of public Economics* 92, no. 10-11 (2008): 1942-1967.
- Bohara, Alok K., Neil J. Mitchell, Mani Nepal, and Nejem Raheem. "Human rights violations, corruption, and the policy of repression." *Policy Studies Journal* 36, no. 1 (2008): 1-18.
- Breen, Richard, Kristian Bernt Karlson, and Anders Holm. "Total, direct, and indirect effects in logit and probit models." *Sociological Methods & Research* 42, no. 2 (2013): 164-191.
- Bueno de Mesquita, Ethan. "The quality of terror." *American journal of political science* 49, no. 3 (2005): 515-530.
- Chenoweth, Erica and Maria J. Stephan. *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press, 2011.
- Fearon, James D., and David D. Laitin. "Ethnicity, insurgency, and civil war." *American political science review* 97, no. 1 (2003): 75-90.
- Fjelde, Hanne, and Indra De Soysa. "Coercion, co-optation, or cooperation? State capacity and the risk of civil war, 1961—2004." *Conflict management and peace science* 26, no. 1 (2009): 5-25.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. "Armed conflict 1946-2001: A new dataset." *Journal of peace research* 39, no. 5 (2002): 615-637.
- Hendrix, Cullen S. "Head for the hills? Rough terrain, state capacity, and civil war onset." *Civil Wars* 13, no. 4 (2011): 345-370.
- Karlson, Kristian Bernt, Anders Holm, and Richard Breen. "Comparing regression coefficients between same-sample nested models using logit and probit: A new method." *Sociological methodology* 42, no. 1 (2012): 286-313.
- Kaufmann, Daniel and Kraay, Aart and Mastruzzi, Massimo. "The Worldwide Governance Indicators: Methodology and Analytical Issues." (September 2010). World Bank Policy Research Working Paper No. 5430. Available at SSRN: <https://ssrn.com/abstract=1682130>

- Kohler, Ulrich, Kristian Bernt Karlson, and Anders Holm. "Comparing coefficients of nested nonlinear probability models." *The Stata Journal* 11, no. 3 (2011): 420-438.
- Marshall, Monty G., Ted Robert Gurr, and Keith Jagers. "Polity IV project: Political regime characteristics and transitions, 1800-2018." *Center for Systemic Peace* (2019).  
<http://www.systemicpeace.org/inscr/p4manualv2018.pdf>
- Mitchell, Neil. *Agents of atrocity: Leaders, followers, and the violation of human rights in civil war*. Springer, 2004.
- Nielsen, Richard A., Michael G. Findley, Zachary S. Davis, Tara Candland, and Daniel L. Nielson. "Foreign aid shocks as a cause of violent armed conflict." *American Journal of Political Science* 55, no. 2 (2011): 219-232.
- Pettersson, Therése, Stina Högladh, and Magnus Öberg. "Organized violence, 1989–2018 and peace agreements." *Journal of Peace Research* (2019): 0022343319856046.
- Powell, Robert. "War as a commitment problem." *International organization* 60, no. 1 (2006): 169-203.
- Rasler, Karen. "Concessions, repression, and political protest in the Iranian revolution." *American Sociological Review* (1996): 132-152.
- Rotberg, Robert I. "Good governance means performance and results." *Governance* 27, no. 3 (2014): 511-518.
- Rotberg, Robert I., ed. *State failure and state weakness in a time of terror*. Brookings Institution Press, 2004.
- Sobek, David. "Masters of their domains: The role of state capacity in civil wars." *Journal of Peace Research* 47, no. 3 (2010): 267-271.
- Walsh, James I., and James A. Piazza. "Why respecting physical integrity rights reduces terrorism." *Comparative Political Studies* 43, no. 5 (2010): 551-577.
- Walter, Barbara F. "Bargaining failures and civil war." *Annual Review of Political Science* 12 (2009): 243-261.

## ***Co-Evolution Modeling***

Victor Asal  
Department of Political Science  
University at Albany, SUNY  
[vasal@albany.edu](mailto:vasal@albany.edu)

R. Karl Rethemeyer  
Department of Public Administration  
University at Albany, SUNY  
[kretheme@albany.edu](mailto:kretheme@albany.edu)

Joseph K. Young  
School of International Service and  
School of Public Affairs  
American University  
[jyoung@american.edu](mailto:jyoung@american.edu)

## Introduction

Previous research has studied the determinants of terrorist lethality ([Asal and Rethemeyer 2008](#); Horowitz and Potter 2014) and terrorist network formation ([Asal, Park, Rethemeyer and Ackerman 2015](#)). However, these studies have relied on cross-sectional data and usually focus on a compressed period (1998-2005). While there have been numerous studies of terrorist organizations in the Middle East, including one study on pursuit and use of CBRN capabilities by organizations in the Middle East ([Asal and Rethemeyer 2009](#)), few of these studies have been both quantitative and longitudinal. To our knowledge no other study has explored quantitatively the *co-evolution* of Violent Nonstate Actor (VNSA) networks and violent behavior on a global scale and across organizations that are classified as terrorist, insurgent, or contentiously political in nature (see Asal, Rethemeyer, Young 2016 for a co-evolution study that focuses on the Middle East and North Africa).

This study provides a preliminary analysis of VNSA lethality and VNSA network/behavior *co-evolution* using the BARK data prepared as part of this project.

## Network and Behavior Co-Evolution in the Middle East: A Preliminary Model

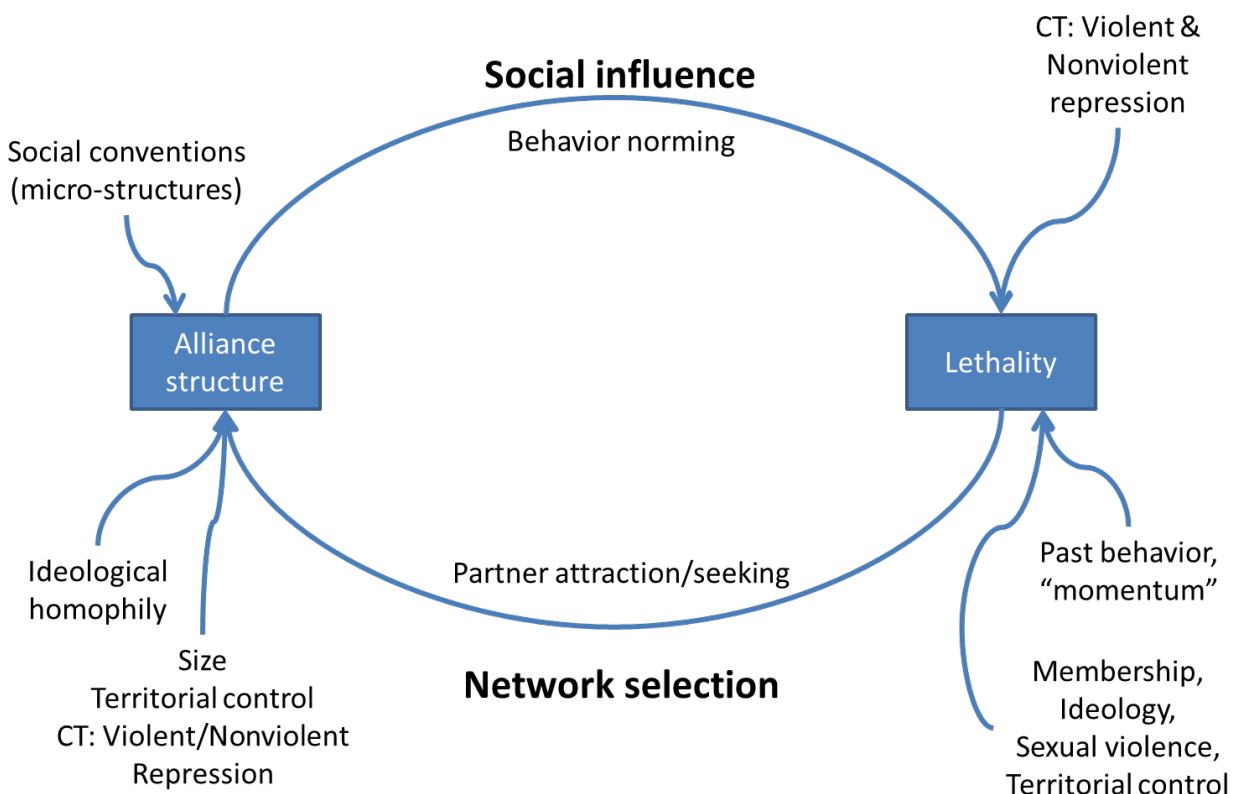
As noted above, previous research has demonstrated that terrorist lethality is influenced by network connections, and network formation is influenced by lethality. However, these findings were generated in two separate studies using a single cross-sectional dataset with highly disparate measures. Ideally, the best way to explore the reciprocal relationship between network formation and lethality is through a simultaneous model that uses longitudinal data. The BARK dataset was created in part to solve the data deficits that inhibited earlier studies. Development of “co-evolution” modeling of network and behavior data by Tom A. B. Snijders and colleagues has resolved the methodological and technical issues with respect to this problem. The results below are, to our knowledge, the first use of this form of analysis to study terrorism or VNSA organizations on a global basis, following on our earlier, regional work – see Asal, Rethemeyer, Young (2016).

## A Brief Theoretical Introduction

Social network analysts have long realized that network formation and behavior are deeply intertwined with one another. However, most analysts have attacked only one side of this feedback loop at a time. Quantitative models of VNSA behavior sometimes include summary measures of network structure – for instance, measures of network centrality – but usually examine only the “influence” channel. That is, these models assume that network position influences behavior. Another group of analysts have sought to understand formation of VNSA networks, their structure, and their evolution. This modeling attacks the “selection” channel, seeking to explain how behavior helps to structure partner selection in networks. Both literatures have created important insights into the behavior of VNSA organizations. However, it is also possible that a simultaneous model may reveal insights that have not previously been discovered. Additionally, stochastic co-evolution models provide a more rigorous methodological foundation for studying these questions (assuming, of course, that both influence and selection operate simultaneously), as the well-known issues of simultaneity and observational dependence that

may bias other approaches to study of behavior within networks are explicitly modeled in this approach.

In Figure 1 we have sketched out a basic co-evolutionary model for VNSA organizational alliance formation and lethality. Co-evolution models allow one to include exogenous factors that affect only network or behavior as well as endogenous influence and selection effects. Our model is greatly influenced by our previous work on lethality ([Asal and Rethemeyer 2008](#)), network formation ([Asal, Park, Rethemeyer and Ackerman 2015](#)), our unpublished work on VNSA network formation and change, and our previous regional co-evolution modeling (Asal, Rethemeyer, Young 2016).



Starting with the alliance side, our previous research found that terrorist networks tend to be:

- Sparse: there is a preference against making connections because connections are costly and dangerous – the new contact could be a counterterrorism operative.
- Closed: there is a preference to make connections such that a friend of a friend is also a friend; closed networks improve the chance to monitor the fidelity of each partner and assure against infiltration.
- Ideologically homophilous: Birds of an ideology feather tend to flock together – and avoid those with differing ideological persuasions.
- Safety-seeking: Organizations prefer to make friends with peers that control territory that can serve as a “safe haven.”

- Capability-seeking: Organizations want friends that have financial, material, or knowledge resources.

However, behavior also seems to drive connections:

- Capability seeking: Organizations that kill more may connect more in order to gather resources to maintain operational tempo.
- Success = popularity: Operationally active organizations become more visible and “popular” because they have demonstrated knowledge and capability.

Thus we expected a “selection” effect from lethality to alliance structure.

With respect to the lethality side of the wheel, our previous research and our concurrent regression modeling of lethality found that VNSA organizations tend to kill when:

- Momentum builds: Killing today helps to explain killing tomorrow.
- Ideology drives decisionmaking: Our research, starting with Asal Rethemeyer (2008), has consistently found that organizations inspired by religious and/or ethnonationalist ideologies kill more.
- Counter-terrorism choices may induce violence: Violent suppression by states may goad organizations into a response in kind.

However, network factors may also help to explain how lethal an organization can be or chooses to be:

- Norming: Organizations may choose to norm their level of killing to the level of killing among their alliance connections.
- Resource gathering: The number of alliance connections may indicate the level of resources from secondary sources that are available to invest in lethal attacks.

So we also expect a social influence effect.

The BARK data for the first time also provides indicators for use of sexual violence. We have included those variables in our modeling also.

## **Modeling**

In order to estimate the model in Figure 1 we used the R-based Simulation Investigation for Empirical Network Analysis (RSIENA) package (Ripley, Snijders, Boda, Voros and Preciado 2019; Snijders 2001; Snijders 2005; Snijders and Duijn 2002; Snijders, Steglich and Schweinberger 2007). RSIENA implements a Stochastic Actor-oriented Model (SAOM) of network and behavior change that is estimated through a Markov Chain Monte Carlo (MCMC) simulation process. For an introduction to RSIENA and the technical aspects of co-evolution modeling, see Ripley, Snijders, Boda, Voros, and Preciado (2019).

In order to account for some gaps in the data and some degree of instability due to variation in reporting intensity in our textual sources across years, we decided to use four-year moving averages for several of our variables. That is, the current year plus three previous years are included in the moving average. Thus while we have 13 years of data, the first three years of data are used to build averages needed to begin modeling with the year 2008. Our model uses 10 years of the BARK data, but models 9 transitions (the year 2008 serves as the initial year for the MCMC analysis process). As is the case for all co-evolution models, our estimations take the 10 yearly alliance networks and the 10 yearly reports of total fatalities from political violence as the network and behavior dependent variables, respectively. The other independent variables include:

- Three ideology indicator variables:
  - Religious ideology
  - Ethnonationalist ideology
  - Separatist ideology
- An indicator whether an organization controlled territory in the last three years or the current year.
- An indicator whether an organization was subject to a violent counterterrorism strategy in the last three years or the current year.
- An indicator whether an organization was subject to a non-violent counterterrorism strategy in the last three years or the current year.
- An count of the number of years in the in the last three years or the current year in which the organization staged “large protests,” which are defined “as demonstrations, rallies, strikes, and/or riots with a total participation of 10,000 or greater” (see BARK codebook).
- An indicator whether an organization used sexual violence in the last three years or the current year.
- A four-year moving average of the organization’s size, where the size variable is a set of ranges:

|   |                                                                              |
|---|------------------------------------------------------------------------------|
| 0 | Unknown (Use when there is no indication of organization size in any source) |
| 1 | 1-9 Members                                                                  |
| 2 | 10-99 Members                                                                |
| 3 | 100-999 Members                                                              |
| 4 | 1,000-9,999 Members                                                          |
| 5 | 10,000-29,999 Members                                                        |
| 6 | 30,000-49,999 Members                                                        |
| 7 | 50,000+ Members                                                              |

- Fatalities from political violence
  - A count of fatalities each year for each organization, which is the sum of (1) battle deaths reported in the Uppsala Conflict Data Program (UCDP) dataset from operations in the context of an insurgency and (2) terrorist deaths inflicted by the organization, as reported in the Global Terrorism Database (GTD)

- A four-year moving average of fatalities, also calculated by adding together battle deaths from UCDP and terrorist deaths from GTD.

Additionally, the program itself calculated variables to indicate:

- Closed triads: The number of times when an organization was involved in a social arrangement when an ally of ally is also an ally.
- The total number of fatalities inflicted by organizations that are tied to the focal organization in the same year. That is, if Organization A has connections to Organizations B, C, and D that killed 200, 100, and 50 people in 2008, respectively, then the value of this variable is 350 for 2008.

The network data for this is model “undirected”. That is, there was no sense of “direction” where Organization A was coded to be allied with Organization B but Organization B was not coded to be allied with Organization A. Relationships were present or absent. Organizations were also allowed to “enter” and “exit” the data. Some organizations did not form until after the beginning of our timeframe, and some dissolved during our timeframe. RSIENA includes specific procedures to properly model network entry and exit. We used the method of “structural zeros” provided in RSIENA (Ripley, Snijders, Boda, Voros and Preciado 2019, 34-36).

Per the conventions with this analysis, the model reported in the next section is considered to have “converged” because the estimated parameters are able to reproduce the observed data over several thousand simulations with high fidelity. Convergence is assessed by the difference between the actual network and behavior statistics and the average of the modeled statistics across the simulations, where the simulations used the estimated coefficients to generate the network and behavioral outcomes via a Monte Carlo process. The program calculates a t-statistic that measures the difference between the observed and average simulated values given the standard deviation of the simulated values. Convergence is when the t-statistic is very small – that is, finds no statistical difference between the observed values in the data and the average in the simulations. The convention is that all t-statistics should be below 0.10 for excellent convergence. For the model reported next, all parameters are below 0.04. An additional measure of convergence is the overall maximum convergence ratio, which should be below 0.25 (see Ripley, Snijders, Boda, Voros and Preciado 2019, 62). For the final model, the overall convergence ratio is 0.14.

## Findings

Tables 1a and 1b contains the model results for both network and behavior dynamics. The rate parameters in both sections are necessary controls but not interpretable (so-called “nuisance parameters”) in the context of this study. The other parameter estimates are similar to logit coefficients: the magnitude is not directly interpretable, but the sign and statistical significance are. The t-statistics reported Tables 1a and 1b are not convergence statistics but traditional t-statistics that indicate whether a coefficient is significantly different from zero given the standard error. A value of 2 or higher is strongly significant; values between 2.00 and 1.66 are weakly significant; values below 1.66 are insignificant.

| Table 1a: Network Dynamics                                                                         |        |           |             |
|----------------------------------------------------------------------------------------------------|--------|-----------|-------------|
|                                                                                                    | Coef.  | Std error | t-statistic |
| Rate, period 1                                                                                     | 7.223  | (0.957)   | .           |
| Rate, period 2                                                                                     | 6.8    | (0.698)   | .           |
| Rate, period 3                                                                                     | 8.17   | (0.818)   | .           |
| Rate, period 4                                                                                     | 8.256  | (0.977)   | .           |
| Rate, period 5                                                                                     | 8.248  | (0.732)   | .           |
| Rate, period 6                                                                                     | 14.018 | (1.419)   | .           |
| Rate, period 7                                                                                     | 8.639  | (0.630)   | .           |
| Rate, period 8                                                                                     | 9.288  | (0.773)   | .           |
| Rate, period 9                                                                                     | 7.503  | (0.600)   | .           |
| Propensity to form connections                                                                     | -3.245 | (0.050)   | -64.433     |
| Network closure (transitive triads)                                                                | 0.397  | (0.012)   | 31.757      |
| Same ethnonationalist ideology (1=yes)                                                             | 0.11   | (0.026)   | 4.213       |
| Same religious ideology (1=yes)                                                                    | 0.291  | (0.031)   | 9.368       |
| Same separatist ideology (1=yes)                                                                   | 0.3    | (0.036)   | 8.266       |
| Similarity in Membership (4 year moving average)                                                   | 0.33   | (0.087)   | 3.816       |
| Propensity to make connections x Membership (4 year moving average)                                | 0.042  | (0.019)   | 2.234       |
| Propensity to make connections x Territory control this year or in any of last 3 years (1=yes)     | 0.294  | (0.062)   | 4.698       |
| Propensity to make connections x Count, years engaged in large protests (this year + last 3 years) | 0.13   | (0.013)   | 10.066      |
| Similar count of years engaged in large protests (this year + last 3 years)                        | -0.158 | (0.035)   | -4.464      |
| Same, used sexual violence this year or in any of last 3 years (1=yes)                             | 0.08   | (0.025)   | 3.215       |
| Same, subject to non-violent repression in this year or any of the last 3 years (1=yes)            | 0.098  | (0.030)   | 3.28        |
| Same, subject to violent repression in this year or any of the last 3 years (1=yes)                | 0.297  | (0.042)   | 7.032       |
| Similar average total fatalities (UCDP + GTD) (4-year moving average)                              | -0.714 | (0.115)   | -6.221      |

## Network dynamics

The dependent network here is alliance connections – 10 yearly 370 x 370 matrices where the “structural zeros” procedure eliminates possible dyads for organizations that are not both in existence in a given year.

The results confirm and extend our previous work. As expected, there is a strong tendency for organizations to *avoid* making connections (see the negative coefficient on *Propensity to form connections*), presumably because it is costly and dangerous as noted above.

However, the modeling does find three strong bases for connections linked to homophily: shared ethnonationalist ideology, shared religious ideology, and/or shared separatist ideology (see, respectively, *Same ethnonationalist ideology*, *Same religious ideology*, and *Same separatist ideology*). Adherents of these ideologies tend to seek relationships with co-adherents substantially more often than adherents to other ideologies. The comparison case is a composite of other, unmodeled ideologies (that is, leftists, rightist, nationalist, separatists, environmentalist, supremacist, vigilantes, and organizations committed to gender exclusion, regime change, or no particular ideology). Also as expected, territorial control is an important factor (see *Propensity to make connections x Territory control*). Because our data is not directional we cannot say whether those that control territory become more popular and garner more “inbound” connection, or whether territory holders become more “outgoing,” possibly because they can more securely make connections knowing that their main base of operations is safe even if the connection proves to be hostile.

Echoing our cross-sectional findings regarding terrorist organizations only, the BARK data suggests that VNSAs make connections based on membership size. Large organizations tend to have more connections (see *Propensity to make connections x Membership*) and to prefer to have connections to organizations of a similar size (see *Similarity in Membership*). This may reflect organizational preference for groups that have relatively similar capability profiles and are able to complement one another as equals, as opposed to relationships where one organization is clearly the senior, more capable partner. For future efforts, the size effect could be further disaggregated by collecting data on various measures of capability as a way to disentangle more thoroughly why organizations prefer partners of similar size.

The BARK data for the first time allows us to look at engagement in non-violent political behavior as a motivator for network relations. Here, we discovered that organizations engaged in large protests often tend to make more connections (see *Propensity to make connections x Count, years engaged in large protests*), but those who protest a lot tend not to connect to others that protest a lot (see *Similar count of years engaged in large protests*). The “similarity” result is interesting because rough 75-85% of all organizations have engaged in a large protest at least once in a given four year period. So organizations may well value relationships with organizations that do not use protest but instead use other forms of political engagement, including violence. Alliances may form in order to help learn new approaches to politics.

The BARK data also makes it possible to examine the effect of sexual violence on alliance formation for the first time. Here, we found no relationship to propensity to form relationships,

but we did find a preference for sexual violence users to connect with each other – and thus also for non-users to avoid each other. The results here suggest a sorting by willingness to use a relatively taboo form of political violence. A future project should focus on exploring the degree to which use of taboo violence leads to subnetwork formation among co-users.

BARK also coded both non-violent and violent repression as a counter-terrorism tactic. While neither form of repression affected overall propensity to make alliances, we found again a tendency for organizations to preferentially ally with other organizations that face similar forms of repression (see both *Same, subject to non-violent repression* and *Same, subject to violent repression*). The effect is much more pronounced for organizations that face violent repression. The upshot is that organizations facing repression tend to seek each other out, presumably to share resources, knowledge, and tactics.

Finally, we also found a clear behavioral dynamic, but one that is somewhat surprising: organizations seek to make connections with organizations that are *dissimilar* in terms of their level of killing (see *Similar average total fatalities (UCDP + GTD)*). Because this finding was unexpected, we tested this result with several different specifications, but all came up statistically significant ( $p < 0$  to at least 4 digits) and relatively large in terms of effect size.

One explanation for this pattern is simply that there are far more organizations in the BARK dataset that have zero or only a few killings in any given year and across the full 13 year period. Precisely 50% of BARK organizations killed no between 2008 and 2017; 60% killed fewer than 10 people; and 70% killed fewer 25 or fewer. This is not anomalous: we have found in several other studies (for instance, Asal & Rethemeyer 2008) that most organizations do not kill at all in a given year and kill infrequently over many years, but a few kill prolifically.

However, the preference for ties between organizations with different levels of killing is also consistent with other research that has found that young, inexperienced, or ineffective organizations seek out prolifically lethal organizations in order to learn. The Lod Airport Massacre by the Japanese Red Army (JRA) in 1972 is one of the most famous examples of this alliance “selection effect”, where the PFLP’s demonstrated competence at killing led the JRA to an alliance to secure training. The training came at a cost: the JRA had to execute the airport attack. Note that the “killing effect” may be contra to our finding regarding the “size” effect, where large organizations seek other large organizations.

These findings clearly suggests that alliance choices are complex, with the various effects interacting with one another. Why might a small organization, which could be seeking support from a large, but reluctant, peer overcome the size effect? It may be due to factors like preference for ideological homophily, triadic closure (for example a common ally of the small organization and the large organization “vouches” for the small organization to the large ally), or relations built to learn new tactics (the large organization may want to learn about using protests as an alternative to violence from the small organization).

## Behavior dynamics

Table 1b contains the behavioral results.

| <b>Table 1b: Behavior Dynamics</b>                                     |        |           |             |
|------------------------------------------------------------------------|--------|-----------|-------------|
|                                                                        | Coef.  | Std error | t-statistic |
| Rate, period 1                                                         | 4.078  | (0.790)   | .           |
| Rate, period 2                                                         | 2.406  | (0.495)   | .           |
| Rate, period 3                                                         | 3.723  | (1.028)   | .           |
| Rate, period 4                                                         | 8.036  | (2.517)   | .           |
| Rate, period 5                                                         | 2.693  | (0.575)   | .           |
| Rate, period 6                                                         | 2.15   | (0.348)   | .           |
| Rate, period 7                                                         | 2.41   | (0.465)   | .           |
| Rate, period 8                                                         | 2.277  | (0.422)   | .           |
| Rate, period 9                                                         | 1.952  | (0.340)   | .           |
| Change over time, linear term                                          | −1.953 | (0.050)   | −39.311     |
| Change over time, quadratic term                                       | 0.467  | (0.017)   | 27.012      |
| Total fatalities inflicted by allies this year                         | 0.022  | (0.007)   | 2.948       |
| Ethnonationalist ideology (1=yes)                                      | 0.344  | (0.082)   | 4.206       |
| Religious ideology (1=yes)                                             | 0.182  | (0.085)   | 2.152       |
| Non-violent repression in this year or any of the last 3 years (1=yes) | −0.209 | (0.096)   | −2.175      |
| Violent repression in this year or any of the last 3 years (1=yes)     | 0.037  | (0.113)   | 0.33        |
| Territory control this year or in any of last 3 years (1=yes)          | −0.156 | (0.101)   | −1.542      |
| Count, years engaged in large protests (this year + last 3 years)      | −0.047 | (0.032)   | −1.458      |
| Used sexual violence this year or in any of last 3 years (1=yes)       | −0.023 | (0.083)   | −0.279      |

The behavioral component of co-evolution models are similar to multinomial logit models, where changes from one level to another are modeled. For this reason the dependent variable, fatalities from political violence, was divided into five categories – see Table 2. There is no hard and fast rule about the best way to divide a continuous variables into categories. Ripley, Snijders, Boda, Voros, and Preciado (2019) suggest that the number not exceed 10. Following past practice, we selected a subdivision that captured the difference between not killing and killing (the transition from category 1 to category 2) and then broke the non-zero fatality organization-years roughly into quartiles:

**Table 2: Levels of Political Killing**

| Category | Min | Max   |
|----------|-----|-------|
| 1        | 0   | 0     |
| 2        | 1   | 4     |
| 3        | 5   | 23    |
| 4        | 24  | 101   |
| 5        | 102 | 50000 |

As with the network dynamics, the rate parameters are treated as “nuisance parameters” and ignored.

The linear and quadratic behavior coefficients (*Change over time, linear term* and *Change over time, quadratic term*, respectively) tell an interesting story. The negative and significant linear term indicates that, on average, most organizations do not kill very much. As noted before, most VNSA organizations never kill. However, the significant quadratic term suggests that once organizations start killing they tend to move to higher and higher levels of killing. In co-evolutionary models a positive coefficient on the quadratic term is usually indicative of an “addictive” behavior. That is, a behavior that feeds on itself: killing increases year to year once an organization first engages in lethal attacks.

The next variable, *Total fatalities inflicted by allies this year*, closes the co-evolutionary loop. This suggests that organizations tend to norm their killing to the total killing of their allied peers in a given year. There are a number of ways to specify this term, including average similarity, average similarity of highly connected peers, controls for isolates, and so forth. We tested all of these: only the yearly total specification is statistically significant. Once we discovered this effect, we also tested to see if the finding was sensitive to inclusion of other exogenous variables in the behavioral component of the model. It was not: the result is robust across several alternative sets of included exogenous behavioral factors is significant.

This finding suggests an accelerator effect on killing from the actions of peers: If my allies are killing a lot, my level of killing will also go up, but it depends on the totality of killing by my allies. So a single, highly lethal ally could drive connected organizations to become more lethal, or a larger group of allies all with some lesser level of individual levels of lethality that are in sum large could have the same effect. Again, the effect is not from an average of lethality across allies, but more about the “exposure” to a relatively high level of lethality from one lethal ally or a group of collectively lethal allies.

The next two variables confirm what several studies (starting with Asal & Rethemeyer 2008) have found: ethnonationalist and religious organizations tend to kill more than organizations inspired by other ideologies (see *Ethnonationalist ideology* and *Religious ideology*). In the BARK data, the effect from ethnonationalist is particularly pronounced.

The final significant effect is the effect of non-violent repression (*Non-violent repression in this year or any of the last 3 years*). This form of repression is generally associated with lower levels of killing. Contrary to other studies we have conducted, this data finds no effect from violent repression (*Violent repression in this year or any of the last 3 years*). Collectively, these two

findings are partially in harmony with recent findings that suggest that non-violent CT tends to reduce violence while violent CT may increase violence. Here, we have not found violent CT increases violence, but there is no evidence it decreases violence, either.

Finally, we included four other effects in the original modeling, three of which are reported in the final table, that were not significant. Contrary to previous findings, we found no relationship between membership (not reported), territorial control, large protests, or use of sexual violence and killing. We removed the size variable because the coefficient was so small in all of the alternative specifications we tried. Of the three variables that were included but found to be insignificant, territorial control is the largest, closest to significant, and most surprising. The coefficient on *Territory control this year or in any of last 3 years* is negative, which is contra to previous findings, starting with Asal & Rethemeyer 2008. The finding on protest is not surprising, as protests and killing may be substitutes for one another. Sexual violence could be a substitute for killing, but the BARK data has absolutely no evidence to support that supposition.

## **Conclusion**

Our analysis of VNSAs in the BARK data from 1998 to 2012 has some important implications for why certain organizations are so lethal, how CT policies impact this lethality, and why some organizations connect and others do not. With respect to alliance network formation, our analysis found several avenues for network formation, including homophily with respect to ideology, preference for closed networks of trusted peers, control of territory, preference for peers of like size, use of sexual violence, repression, and seeking of peers with different protest behaviors. We also identify a key behavioral influence on network formation: those who kill a lot tend to connect to those who do not and vice versa.

With respect to lethality, most organizations do not kill. Indeed, 50% of BARK VNSAs never killed between 2008 and 2017. Nevertheless, we find a set of important predictors of terrorist lethality, including ideology, and non-violent repression. Most interestingly, we found an influence effect, where organizations that are connected to lethal peers tend to increase their killing as the total killing in their alliance subnetwork increases. Finally, the model suggests that terrorist lethality has an “addictive” dynamic: once killing commences, organizations tend to move to higher levels of killing over time.

## References

- Asal, V., H. H. Park, R. K. Rethemeyer and G. Ackerman. 2015. "With Friends Like These....: Why Terrorist Organizations Ally." *International Public Management Journal*.
- Asal, V. H. and R. K. Rethemeyer. 2008. "The Nature of the Beast: Organizational Structures and the Lethality of Terrorist Attacks." *Journal of Politics* 70(2): 437-449.
- Asal, V. H. and R. K. Rethemeyer. 2009. "Islamist Use and Pursuit of CBRN Terrorism." in G. Ackerman and J. Tamsett, eds., *Jihadists & Weapons of Mass Destruction: A Growing Threat*, Boca Raton, FL: CRC Press.
- Asal, V., R. K. Rethemeyer, and M. Young. (2016). "An Analysis of Violent Nonstate Actor Organizational Lethality and Network Co-Evolution in the Middle East and North Africa." College Park: National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland.  
[https://www.start.umd.edu/pubs/START\\_I%26A\\_BAAD\\_AssessingMiddleEasternVNSALethality\\_May2016.pdf](https://www.start.umd.edu/pubs/START_I%26A_BAAD_AssessingMiddleEasternVNSALethality_May2016.pdf)
- Horowitz, Michael C., and Philip BK Potter. "Allying to kill terrorist intergroup cooperation and the consequences for lethality." *Journal of Conflict Resolution* 58.2 (2014): 199-225.
- Ripley, R. M., T. A. B. Snijders, Z. Boda, A. Voros and P. Preciado. 2019. *Manual for SIENA version 4.0 (version May 21, 2019)*. Oxford: University of Oxford, Department of Statistics; Nuffield College. <http://www.stats.ox.ac.uk/~snijders/siena/>.
- Snijders, T. A. B. 2001. "The Statistical Evaluation of Social Network Dynamics." in M. E. Sobel and M. P. Becker, eds., *Sociological Methodology 2001*, Boston: Basil Blackwell.
- Snijders, T. A. B. 2005. "Models for longitudinal network data." Pp. 215-247 in P. J. Carrington, J. Scott, and S. Wasserman, eds., *Models and methods in social network analysis*, New York: Cambridge University Press.
- Snijders, T. A. B. and M. A. J. Duijn. 2002. "Conditional maximum likelihood estimation under various specifications of exponential random graph models." in J. Hagberg, ed., *Contributions to social network analysis, information theory, and other topics in statistics: A Festschrift in honour of Ove Frank*, Stockholm: University of Stockholm, Department of Statistics.
- Snijders, T. A. B., C. E. G. Steglich and M. Schweinberger. 2007. "Modeling the co-evolution of networks and behavior." Pp. 41-72 in K. v. Montfort, H. Oud, and A. Satorra, eds., *Longitudinal models in the behavioral and related sciences*, Mahwah, NJ: Lawrence Erlbaum.

***Appendix I: The Organization that Does(n't) BARK:  
Organizations and the Decision for Political Violence Codebook***

## Contact Information

### **Victor Asal, PhD**

Professor of Political Science  
Rockefeller College, University at Albany – SUNY  
Milne Hall 217  
135 Western Avenue  
Albany, NY 12203  
(518) 591-8729  
[vasal@albany.edu](mailto:vasal@albany.edu)

### **R. Karl Rethemeyer, PhD**

Dean & Professor  
Rockefeller College, University at Albany – SUNY  
102B Milne Hall  
135 Western Avenue  
Albany, NY 12203  
(518) 442-5283  
[kretheme@albany.edu](mailto:kretheme@albany.edu)

### **Michael Young, PhD**

Assistant Professor  
College of Emergency Preparedness, Homeland Security & Cybersecurity  
University at Albany – SUNY  
342 Draper Hall  
135 Western Avenue  
Albany, NY 12203  
(518) 442-5783  
[myoung4@albany.edu](mailto:myoung4@albany.edu)

## The Organization that Does(n't) BARK Project History

The BARK project combines several variables from existing datasets to track the behavior of violent and non-violent non-state organizations. Unlike previous projects, the primary method of data collection is via automated coding. This coding is completed in partnership with Dr. Michael Young and Michelle Young of Social Science Automation Inc., a company that specializes in developing tools for manipulating and analyzing unstructured text. More information about the company and their work can be found at: <http://socialscience.net/>

BARK includes quantitative variables from both the Minorities at Risk Organizational Behavior (MAROB)<sup>2</sup> and Big Allied & Dangerous (BAAD)<sup>3</sup> projects.

The Minorities at Risk Organizational Behavior (MAROB) dataset is one of the only efforts that collects yearly data on both violent and nonviolent as well as non-contentious political organizations. The purpose of this project is to answer fundamental questions focusing on the identification of those factors that motivate some members of ethnic minorities to become radicalized, to form activist organizations, and to move from conventional means of politics and protest into violence and terrorism.

The Big Allied & Dangerous (BAAD) database was created by Dr. Victor Asal and Dr. R. Karl Rethemeyer and tracks the organizational and network behavior of terrorist and insurgent organizations. It was devised to gain increased understanding of how violent non-state actors pursue their objectives individually and collectively through networks. The dataset is made up of three major components: organizational variables, social network data, and the terrorist organization (TORG) identification system.

The BARK dataset contains several variables from both the BAAD and MAROB projects. Text examples were extracted from BAAD and MAROB in order to create a coding scheme for individual variables. This scheme was used by Social Science Automation, Inc. to generate data from selected publications for the years 2005 to 2017. A complete list of these publications can be found in Appendix I.

---

<sup>2</sup> Asal, Victor, Amy Pate and Jonathan Wilkenfeld. 2008. Minorities at Risk Organizational Behavior Data and Codebook Version 9/2008 online: <http://www.mar.umd.edu/data.asp>

<sup>3</sup> Asal, Victor H. and R. Karl Rethemeyer. (2015). Big Allied and Dangerous Dataset Version 2. Available for browsing at <http://www.start.umd.edu/baad/database>

## The Organization that Does(n't) Bark Inclusion Criteria

The BARK project combines variables from two datasets in order to track the behavior of violent and non-violent non-state organizations.

**Big Allied And Dangerous (BAAD) Complete Universe**, which includes:

1. All violent non-state organizations that perpetrated at least one terrorist “event” that is recorded in the Global Terrorism Database (GTD) on or after January 1, 1998 through December 31, 2012, for which there is sufficient information to record the organization’s founding year, home base country, (in almost all cases) primary ideology, and at least some of the other organizational variables defined below. The definition of terrorist attack used by the GTD may be found at <http://www.start-dev.umd.edu/gtd/using-gtd/>. In some cases organizations included in GTD are too amorphous to be considered “organizations” or too poorly covered in the new media to generate even the most basic of information for coding.
2. All violent non-state organizations that participated in a battle in which at least 25 battle-related deaths occurred, as defined and recorded by the Uppsala Conflict Data Program (UCDP) on or after January 1, 1998 through December 31, 2012, for which there is sufficient information to record the organization’s founding year, home base country, (in almost all cases) primary ideology, and at least some of the other organizational variables defined below. The definition of battle death is found at <http://www.pcr.uu.se/research/ucdp/definitions/>. Battle deaths usually refer to killing of armed military, paramilitary, or police forces.
3. All organizations that appear in the Profiles of Incidents Involving CBRN Use by Non-state Actors (POICN) database for which there is sufficient information to record the organization’s founding year, home base country, (in almost all cases) primary ideology, and at least some of the other organizational variables defined below. POICN is a dataset constructed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) that contains over 400 uses and attempted uses of weapons of mass destruction. For more information on POICN, please contact Gary Ackerman ([ackerman@umd.edu](mailto:ackerman@umd.edu)).

In some cases “organizations” identified in these primary sources were either:

- too amorphous (“Chechens”; “Palestinians”),
- too ephemeral (did not exist for long and/or were poorly covered in the media), and/or
- lacked a clear organizational existence, which we defined as having some clear indication that the entity had corporate property (as opposed to property held by individual members) and some boundary that defined the difference between members and non-members

to adequately code. The final universe in this release includes 514 violent non-state organizations.

### **Minorities at Risk Organizational Behavior (MAROB) dataset:**

Organizations are coded for MAROB if they meet all of the following inclusion criteria:

- The organization makes explicit claims to represent the interests of one or more ethnic groups and/or the organization's members are primarily members of a specific ethnic minority.
- The organization is political in its goals and activities.
- The organization is active at a regional and/or national level.
- The organization was not created by a government.
- The organization is active for at least three consecutive years between 1980 and 2006.
- The organization does NOT act as an umbrella organization (coalition/alliances). Instead, member organizations are coded.

Organizations were selected on the basis of their prominence. This was operationalized in the following manner: The first year that an organization is mentioned in a source as being active, it is put on a "watchlist" for potential inclusion. Once the organization is mentioned in sources for three consecutive years, it is included in the dataset, coded from the first year of the three consecutive years. If an organization included in the dataset disappears from source material for five consecutive years, it is no longer coded for following years. If after that time, it is again mentioned for three consecutive years, it is again included but as a separate organization. Organizations are not coded from their founding date as they are often not active in manner that would be reported. Many organizations existed prior to their inclusion in the dataset but were not prominent. Often these organizations were not large enough to attract attention or had no explicit political goals until the first year of their inclusion. Additionally, some of the organizations are part of umbrella organizations that may have been active primarily under the name of the umbrella organization. Action by the organization itself may not be verified in some cases and therefore not included.

### **Umbrella Organizations:**

For the purposes of MAROB, umbrella organizations will not be coded, but rather the individual organizations that make up the umbrella will be coded separately. If there are actions attributed to the umbrella organization only and it is not clear which individual organizations took part, the action will be attributed to the largest/most significant organization under the umbrella. For example, PLO actions that cannot be connected to specific organizations will be attributed to Fatah.

## Organization of the Dataset

The Organization that Does(n't) Bark Release 1.0 dataset contains four separate Excel file datasets for organizational variables and networking relationships:

**BARK\_BAAD Universe:** This dataset contains only those organizations meet the criteria for inclusion into the BAAD Universe.<sup>4</sup>

**BARK\_MAROB MENA:** This dataset contains only those organizations that meet the criteria for the MAROB dataset and have their homebase in the Middle East and North Africa (MENA) region.<sup>5</sup>

**BARK\_MAROB EE:** This dataset contains only those organizations that meet the criteria for the MAROB dataset and have their homebase in the Eastern European (EE) region.<sup>6</sup>

Each dataset is distributed as an Excel workbook with three tabs:

1. Organization List:

This tab contains a list of names for the violent non-state organizations contained in the database and associated ID number.

In the BAAD BARK\_UNIVERSE dataset, the organization's primary name is based on BAAD's Terrorist Organizations (TORG) master list, which includes more than 2,700 primary entities and more than 3,500 aliases and alternate spellings. The name is associated with a unique numeric identifier that represents the primary entity and can be used to look up aliases within BAAD's Terrorist Organizations master list.

In the BARK\_MAROB datasets, the organization's primary name is based on the MAROB master group list. The name is associated with a unique ID number, pulled from the existing MAROB dataset.

---

<sup>4</sup> Myanmar National Democratic Alliance Army (TORG 2585) and United Democratic Liberation Army (Udla) (TORG 2534) are not included in the BAAD\_BARK Universe dataset

<sup>5</sup> Iraqi National Alliance (MAROBID 6450702) is not included in the BAAD\_MAROB MENA dataset

<sup>6</sup> Social Democratic Party of Kosovo (Beqiri) (MAROBID 3450116) is not included in the BAAD\_MAROB EE dataset

## 2. State List:

This tab contains a list of the names for the states found in the database and the associated country code from the following lists:

- Correlates of War (COW)
- International Organization for Standardization (ISO) Alpha-3 Codes
- International Organization for Standardization (ISO) Numeric Codes

## 3. Organizational Variables

This tab contains the organizational variables for all organizations in the dataset. The definition and coding for each organizational variable is found in the next section.

**BARK\_Network:** This dataset contains the network relationships between organizations in the BARK Universe.

The dataset is distributed as an Excel workbook with two tab:

### 1. Organization List:

This tab contains a list of names for the violent non-state organizations contained in the database and associated ID numbers.

### 2. Network:

This tab contains a dyadic list of all organization-organization relationships coded in BARK. This data is a one-mode, multi-relational dataset, where each dyad is defined by the sender, receiver, and year in which the relationship was known to exist.

## Organizational Variables

### Organization Identification

1. ORG (Primary Name of the Organization)

The organization's PRIMARY group name. In BARK\_BAAD Universe, this name is based on BAAD's Terrorist Organizations (TORG) master list. In the BARK\_MAROB datasets, the organization name is based on the MAROB master group list.

2. SUPERID (ID Attributed to Each Organization)

This is the unique numerical identifier for the organization that is input from the BARK group master list.

3. YEAR (Year of Data)

Year the source article was published.

4. BAAD\_ORG (BAAD Organization)

Is this organization part of the BAAD Universe dataset? That is, has it (1) perpetrated at least one terrorist "event" that is recorded in the Global Terrorism Database (GTD) on or after January 1, 1998 through December 31, 2012, (2) participated in a battle in which at least 25 battle-related deaths occurred, as defined and recorded by the Uppsala Conflict Data Program (UCDP) on or after January 1, 1998 through December 31, 2012, or (3) appear in the Profiles of Incidents Involving CBRN Use by Non-state Actors.

0 No

1 Yes

5. MAROB\_ORG (MAROB Organization)

Is this organization part of MAROB dataset? That is, does the organization (1) make explicit claims to represent the interests of one or more ethnic groups and/or the organization's members are primarily members of a specific ethnic minority, (2) is political in its goals and activities, (3) is active at a regional and/or national level, (4) was not created by a government and (5) is active for at least three consecutive years between 1980 and 2006.

0 No

1 Yes

6. HBASE (Organizational Homepage)

Country where the organization is based (textual). This can be thought of as the "main" base/area of operations, as many violent non-state organizations operate in multiple countries. The variable was coded from the existing data in BAAD or MAROB, if the VNSA actor was in both datasets, the BAAD homepage was used.

7. HBASE\_INF (Organizational Homebase Inference)

This variable indicates if homebase country was inferred based on other years of coding or if we found direct evidence for that year. A code of 1 means the homebase is inferred and 0 indicates the homebase was coded directly from an existing BAAD or MAROB dataset for that year. Homebase is inferred from the closest year of data available.

8. HB\_CC CODE (Organizational Homebase– COW)

Numeric country code from the Correlates of War (COW) country identification system.

9. HB\_ISO (Organizational Homebase – ISO Alpha-3 Code)

Three-letter homebase country codes available from the International Organization for Standardization (ISO) accessible at: [http://www.iso.org/iso/country\\_codes.htm](http://www.iso.org/iso/country_codes.htm).

10. HB\_ISO\_CC (Organizational Homebase – ISO Numeric Country Code)

Three digit homebase country codes available from the International Organization for Standardization (ISO) accessible at: [http://www.iso.org/iso/country\\_codes.htm](http://www.iso.org/iso/country_codes.htm).

11. ISO\_REC

This variable was added with the following two changes from HB.ISO.CC: Palestinian territories coded as 376 and Serbia and Montenegro is coded as 10891 to aid with merger of data variously coded across this time period as Yugoslavia, Serbia, Montenegro, and Serbia and Montenegro.

12. GDP\_PER\_CAP\_UN

Yearly GDP per capita information based on the homebase country from the United Nations data. Information was downloaded on February 18, 2019. Available from:  
<http://data.un.org/Data.aspx?d=SNAAMA&f=grID%3A101%3BcurrID%3AUSD%3BpcFlag%3A1>

NOTES: The GDP per capita information for the former counties of Serbia and Montenegro and Yugoslavia coded with the GDP from Serbia.

13. TORG (Violent Non-state Organizational Identifier code) or MAROBID

In the BARK\_BAAD Universe dataset the TORG is the unique numeric identifier that represents the primary entity and can be used to look up aliases within BAAD's Terrorist Organizations master list. The MAROBID in the BARK\_MAROB datasets is unique numerical identifier for groups belonging to the MAROB universe.

#### 14. TORG\_YEAR or MAROBID\_YEAR

Unique identifier for an organization-year. Concatenation of the TORG ID number OR MAROBID number with an underscore (" \_ ") between the ID and year of data reported in the row.

## **Organizational Ideology**

Ideology is the guiding principle for the organization and the closest thing to an organization identity. Our system coded 10 possible ideological motivators. Organizations may adhere to multiple ideologies at once (so-called “compound ideologies”). Ideologies may also change over time, but most organizations are relatively invariant with respect to ideology over time.

Each of the following variables may take only two values:

- 0 Does not adhere to ideology
- 1 Adheres to ideology

### **15. LEFT (Leftist Organization)**

The organization promotes economically leftist policies such as redistribution of wealth by the government and nationalization of industry. Most that fall into this category are communist and socialist organizations and their variants (Marxist, Leninist, Maoist). Examples include the New People’s Army (NPA), the Red Brigades and the Revolutionary Armed Forces of Columbia (FARC).

### **16. LEFT\_MNTINS (Leftist Ideology Mentions)**

A count of how often a leftist ideology was identified for this group in this year.

### **17. RGHT (Rightist Organization)**

Organizations that promote economically rightist policies such as extreme laissez-faire policies, small government, or policies that favor status quo economic elites. Often rightist organizations are formed as a reaction to leftist organizations, such as many anti-communist organizations in Central and South America. Examples include United Self-Defense Forces of Columbia (AUC) and Mountaineer Militia.

### **18. RGHT\_MNTINS (Rightist Ideology Mentions)**

A count of how often a rightist ideology was identified for this group in this year.

### **19. RELI (Religious Organization)**

Organizations guided by some form of religious principles. They may seek to incorporate religious policies into public life or exist to protect a distinct religious organization. Examples include al-Qa’ida, Jemaah Islamiya, Kach, and the Lord’s Resistance Army (LRA).

20. RELI\_MNTINS (Religious Ideology Mentions)

A count of how often a religious ideology was identified for this group in this year.

21. ETHN (Ethnic Organization)

Organizations that represent certain ethnic groups and advocate for the rights or expansion of rights of that ethnic group. Examples include Fatah, Basque Fatherland and Freedom (ETA), and the Kurdistan Workers' Party (PKK).

22. ETHN\_MNTINS (Ethnic Ideology Mentions)

A count of how often an ethnic ideology was identified for this group in this year.

23. SEPA (Separatist Organization)

Organizations that advocate for the autonomy, independence, or annexation of a certain geographical area for a represented group. Often this ideology overlaps with religious or ethnic ideologies. Examples include the Irish Republican Army (IRA), Basque Fatherland and Freedom (ETA), and the Liberation Tigers of Tamil Eelam (LTTE).

24. SEPA\_MNTINS (Separatist Ideology Mentions)

A count of how often a separatist ideology was identified for this group in this year.

25. ENVI (Environmentalist Organization)

Organizations that advocate for policies that favor the conservation of the environment or wildlife. Examples include the Earth Liberation Front (ELF), Animal Liberation Front (ALF), and Oxford Arson Group.

26. ENVI\_MNTINS (Environmentalist Ideology Mentions)

A count of how often an environmentalist ideology was identified for this group in this year.

27. SUPR (Supremacist Organization)

Organizations that advocate for the inherent supremacy of a specific race over other supposedly inferior races. Some extreme nationalist organizations reach this level as well. Examples include the Ku Klux Klan, Aryan Nations, and Russian National Unity.

28. SUPR\_MNTINS (Supremacist Ideology Mentions)

A count of how often a supremacist ideology was identified for this group in this year.

29. VIGL (Vigilante Organization)

Organizations that hold a strong anti-crime goal, often speaking out against the government for not properly dealing with crime or attacking corrupt police. They often attempt to act as a neighborhood watch or an illegitimate police force to fill this perceived void. An example is People Against Gangsterism and Drugs (PAGAD).

30. VIGL\_MNTINS (Vigilante Ideology Mentions)

A count of how often a vigilante ideology was identified for this group in this year.

31. REGCO (Regime Change)

Organizations whose goal is to change the leadership of the country. This includes revolutionary organizations such as the Free Syrian Army.

32. REGCO\_MNTINS (Regime Change Ideology Mentions)

A count of how often a regime change ideology was identified for this group in this year.

33. NATILST (Nationalist)

Organizations that proclaim allegiance to a particular country and/or advocates national independence or a strong national government.

34. NATILST\_MNTINS (Nationalist Ideology Mentions)

A count of how often a nationalist ideology was identified for this group in this year.

35. GENDEXC (Gender Exclusiveness)

Organizations who advocate the exclusion of women from public life.

### 36. GENDEXC\_MNTINS (Gender Exclusiveness)

A count of how often a gender exclusiveness ideology was identified for this group in this year.

## **Organizational Characteristics**

This set of variables provides key information on the demographics and internal structure of organizations included in the dataset.

### **37. MINSIZE (Minimum Organizational Size)**

This is a categorical variable that indicates the lowest estimate of the approximate number of members in the organization using a base-10 logarithmic scale. If an organizational strength estimate spans across two or more size intervals than the lowest size category is coded (Ex. A strength estimate of 50-100 results in a size code of “2”).

- 0 Unknown (Use when there is no indication of organization size in any source)
- 1 1-9 Members
- 2 10-99 Members
- 3 100-999 Members
- 4 1,000-9,999 Members
- 5 10,000-29,999 Members
- 6 30,000-49,999 Members
- 7 50,000+ Members

### **38. MAXSIZE (Maximum Organizational Size)**

This is a categorical variable that indicates the highest estimate of the approximate number of members in the organization using a base-10 logarithmic scale. If an organizational strength estimate spans across two or more size intervals than the lowest size category is coded (Ex. A strength estimate of 50-100 results in a size code of “2”).

- 0 Unknown (Use when there is no indication of organization size in any source)
- 1 1-9 Members
- 2 10-99 Members
- 3 100-999 Members
- 4 1,000-9,999 Members
- 5 10,000-29,999 Members
- 6 30,000-49,999 Members
- 7 50,000+ Members

### **39. AVGSIZE (Average Organizational Size)**

This variable indicates the average all of the estimates of the approximate number of members in the organization using a base-10 logarithmic scale.

- 0 Unknown (Use when there is no indication of organization size in any source)
- 1 1-9 Members
- 2 10-99 Members
- 3 100-999 Members
- 4 1,000-9,999 Members
- 5 10,000-29,999 Members
- 6 30,000-49,999 Members
- 7 50,000+ Members

#### 40. SIZE\_MNTINS (Size Mentions)

A count of how often a size estimate was identified for this group in this year.

#### 41. TERRCNTRL (Territorial Control)

We define territorial control as existing when the organization is able to control movement into, out of, or within a given territory. In some instances organizations will perform functions or provide services, similar to that of a legitimate government. Territory may be controlled by threat or use of force, or if the government grants the organization the authority to do so. The territory must be a substantial area (city, region, etc.) and not just an organization occupying a building or a city block. Thus this excludes military bases and checkpoints. Though bases may indicate that the organization does have control over territory, they may also be covert and thus not exercising social control over civilians in the area of the base. Additionally, the organization must have control over the territory for more than a few days. For instance, seizure and loss of a town in the space of a week does not constitute territorial control.

- 0 Does not control territory
- 1 Controls territory

Inference rule: all -99s were coded as 0 in the data.

#### 42. TERRCNTRL\_MNTINS (Territorial Control Mentions)

A count of how often an instance of territorial control was identified for this group in this year.

#### 43. SEXUL\_VIOL

Does the organization engage in sexual violence? This consists of rape defined as “the coerced (under physical force or threat of physical force against the victim or a third person) penetration of the anus or vagina by the penis or another object, or of the mouth by the penis” (Wood, 2006). This also includes acts that constitute “sexual violence as ‘a broader category that includes rape, coerced undressing and non-penetrating sexual assault’ as well as a variety of other forms of violence”.

- 0 Group does not engage in sexual violence
- 1 Group engages in sexual violence

#### 44. SEXUL\_VIOL\_MNTINS (Territorial Control Mentions)

A count of how often an instance of sexual violence was identified for this group in this year.

## **Protest and Demonstration**

This set of variables records information on the protest activities of organizations, including size and total duration of the demonstrations.

### 45. PROTEST (Protest)

Was any type of protest staged by the organization? This variable includes exclusively instances where the protest type was unknown.

- 0 No
- 1 Yes

### 46. PROTEST\_MNTINS (Protest Mentions)

A count of how often an instance of a protest with an unknown type was identified for this group in this year.

### 47. PROTEST\_DAY (Protest Days)

The total duration of protests with an unknown type, counted in days.

### 48. SYM\_RESIS (Symbolic Resistance)

Was symbolic resistance used by the organization? This includes scattered acts of symbolic resistance (e.g. sit-ins, blockage of traffic, sabotage, symbolic destruction of property), political organizing activity on a substantial scale or mobilization for autonomy/secession by a minority-controlled regional government.

- 0 No
- 1 Yes

### 49. SYM\_RESIS\_MNTINS (Symbolic Resistance Mentions)

A count of how often an instance of symbolic resistance was identified for this group in this year.

### 50. SYM\_RESIS\_DAY (Symbolic Resistance Days)

The total duration of the acts of symbolic resistance, counted in days.

51. SMLL\_DEM (Small Demonstration)

Was a small demonstration staged by the organization this year? This is defined as a few demonstrations, rallies, strikes, and/or riots with total participation of less than 1000.

- 0 No
- 1 Yes

52. SMLL\_DEM\_MNTINS (Small Demonstration Mentions)

A count of how often a small demonstration was identified for this group in this year.

53. SMLL\_DEM\_DAY (Small Demonstration Days)

The total duration of the small demonstrations, counted in days.

54. MED\_DEM (Medium Demonstration)

Was a medium demonstration staged by the organization this year? This is defined as demonstrations, rallies, strikes, and/or riots with a total participation of 1,000 to 9,999.

- 0 No
- 1 Yes

55. MED\_DEM\_MNTINS (Medium Demonstration Mentions)

A count of how often a medium demonstration was identified for this group in this year.

56. MED\_DEM\_DAY (Medium Demonstration Days)

The total duration of the medium demonstrations, counted in days.

57. LRG\_DEM (Large Demonstration)

Was a large demonstration staged by the organization this year? This is defined as demonstrations, rallies, strikes, and/or riots with a total participation of 10,000 or greater.

- 0 No
- 1 Yes

58. LRG\_DEM\_MNTINS (Large Demonstration Mentions)

A count of how often a large demonstration was identified for this group in this year.

59. LRG\_DEM\_DAY (Large Demonstration Days)

The total duration of the large demonstrations, counted in days.

## **Organizational Exposure to Counterterrorism Efforts**

This set of variables records information on sovereign state activities used to suppress the operations of the violent non-state organization. The states involved may be the “homebase” country or other countries.

### **60. ORG\_NEGOT (Organizational Negotiations)**

This is a categorical variable that indicates to what degree the organization has succeeded in obtaining government agreement in the year coded (code is based on agreement, not implementation)?

- 0 No negotiation with government sought
- 1 State refuses to negotiate with the organization
- 2 State has negotiated with the organization in this year

Inference rule: all -99s were coded as 0 in the data.

### **61. ORG\_NEGOT\_MNTINS (Organizational Negotiations Mentions)**

A count of how often organizational negotiation was identified for this group in this year.

### **62. VIOL\_REPRESS (Violent State Repression)**

Does the state engage in violent repression against the organization?

- 0 State is not using violent repression against the organization
- 1 State is using violent repression against the organization

### **63. VIOL\_REPRESS\_MNTINS (Violent State Repression Mentions)**

A count of how often an instance of violent repression was identified for this group in this year.

### **64. NONVIOL\_REPRESS (Non-Violent State Repression)**

Does the state engage in non-violent action against an organization? This includes negative sanctions levied by government agents, arrests, detentions, raids on offices, infiltration and surveillance, confiscation of property, and nonlethal force by government agents against the organization.

- 0 State is not using non-violent repression against the organization
- 1 State is using non-violent repression against the organization

65. NONVIOL\_REPRESS\_MNTINS (Non-Violent State Repression Mentions)

A count of how often an instance of non-violent repression was identified for this group in this year.

## **Organizational Engagement in Violence**

### **66. INCIDENTS (Incidents, per GTD)**

The number of terrorist attacks the organization perpetrated in this year; derived from the GTD.

### **67. FATALITIES (Fatalities, per GTD)**

The number of fatalities the organization perpetrated through terrorist attacks this year; derived from the GTD.

### **68. UCDP.BD (Uppsala Conflict Data Program Battle Death count)**

The number of deaths inflicted in battle between the organization and state this year; derived from the UCDP Battle Deaths dataset.

## **Nature of Organizational Terrorist Attack**

This section includes variables derived from the GTD terrorism event data.

### **69. PM\_INCID (Police and Military Incidents, per GTD)**

The number of terrorist attacks the organization perpetrated each year against military and police targets. This does not exclude attacks where civilian, terrorists/non-state militias, or violent political parties were also targeted.

### **70. NONPM\_INCID (Non Police and Military Incidents, per GTD)**

The number of terrorist attacks the organization perpetrated each year against civilian targets. This excludes attacks where military, police, terrorists/non-state militias, or violent political parties were targeted as well.

### **71. PM\_FATL (Police and Military Fatalities, per GTD)**

The number of fatalities the organization perpetrated each year against military and police targets through terrorist attacks.

### **72. NONPM\_FATL (Non Police and Military Fatalities, per GTD)**

The number fatalities the organization perpetrated through terrorist attacks each year excluding attacks where military, police, terrorists/non-state militias, or violent political parties were also targeted.

## Relational Variables

The BARK relational data is dyadic, multi-relational, and encompasses one-mode organizational data.

### Dyadic Organizational Variables

#### 73. SNDORG (Sending Organization Name)

Name of sending organization. In BARK\_BAAD Universe, this name is based on BAAD's Terrorist Organizations (TORG) master list. In the BARK\_MAROB datasets, the organization name is based on the MAROB master group list.

#### 74. SID\_SUPERID (Sending Organization SUPERID)

SUPERID number of sending organization.

#### 75. SN\_HB\_ISO (Sending Organization Homebase ISO Alpha-3 Code)

The three letter code representing the homebase country of the sending organization. Codes available available from the International Organization for Standardization (ISO) accessible at: [http://www.iso.org/iso/country\\_codes.htm](http://www.iso.org/iso/country_codes.htm).

#### 76. SN\_HB\_ISO\_CC (Sending Organization Homebase ISO Numeric Country Code)

Three digit ISO code representing the homebase country of sending organization.

#### 77. RCVORG (Receiving Organization Name)

Name of receiving organization. In BARK\_BAAD Universe, this name is based on BAAD's Terrorist Organizations (TORG) master list. In the BARK\_MAROB datasets, the organization name is based on the MAROB master group list.

#### 78. RID\_SUPERID (Receiving `Organization SUPERID)

SUPERID number of receiving organization.

#### 79. RC\_HB\_ISO (Receiving Organization Homebase ISO Alpha-3 Code)

The three letter code representing the homebase country of the receiving organization. Codes available from the International Organization for Standardization (ISO) accessible at: [http://www.iso.org/iso/country\\_codes.htm](http://www.iso.org/iso/country_codes.htm).

80. RC\_HB\_ISO\_CC (Receiving Organization Homebase ISO Numeric Country Code)

Three digit ISO code representing the homebase country of receiving organization.

81. RC\_HB\_ISO\_CC (Receiving Organization Homebase ISO Numeric Country Code)

Three digit ISO code representing the homebase country of receiving organization.

82. YEAR

Year the source article was published.

83. AFFLIATN (Affiliation Type)

Our coding scheme includes 2 possible types, for organization-organization relationships.

**Interorganizational Connections**

It is possible to have several different types of interorganizational connections between the same two organizations in one year. For example, Hamas and Fatah are rivals, but in one year they might also work together (ally).

6 Ally

Organizations are known to cooperate operationally, share resources, or in some other means actively provide assistance to one another. Each side of the potential alliance is coded separately. The data is not inherently symmetric in its raw form.

8 Rival

The organizations compete for the same object or goal and are competitors for the same constituents. Rivals seek to dispute another's preeminence or superiority. The most prominent example of a rivalrous relationship is that between Hamas and Fatah.

Each side of a rivalry is coded separately. The data is not inherently symmetric in its raw form.

84. FREQ (Frequency)

A count of how often the recorded relationship was identified in this year.

### Publications Used in the BARK dataset

- ABC News Transcripts
- Agence France Presse - English
- Agencia de Informacao de Mocambique (English)
- Arusha Times (Arusha)
- Cape Argus (Cape Town)
- Concord Times (Freetown)
- Cameroon Tribune (Yaounde) English
- Daily Champion (Lagos)
- Daily Independent (Lagos)
- African Manager (English)
- Africa Newswire
- Africa News<sup>7</sup>
- The Independent (Kampala, Uganda)
- AllAfrica Newspapers (English)
- The Age (Melbourne, Australia) - Online
- Afghan Islamic Press
- Al-Akhbar English
- Algeria Press Service
- Al Jazeera Africa
- Al Jazeera - English
- AmmanNet (English)
- Asian News International (ANI)
- Associated Press International
- African Press Organization
- Associated Press Online
- Al-Arab (English)
- AlArabiya.net
- Arab News
- ASEAN Tribune
- South Asia Monitor
- Assam Tribune (India)
- Afternoon Voice (India)
- Azerbaijan News Gazette
- Bahrain News Agency (BNA)
- Daily Balochistan Express
- Baltic News Service

---

<sup>7</sup> Approximately half of these documents were processed

- Bahrain News Gazette
- Bayane Al Yaoume
- Bangor Daily News (Maine)
- Bharat Chronicle
- Bihar Times (India)
- Birmingham Post
- Birzha.ru-articles
- Birzha-Federal and World News
- Birzha.ru-news Nizhnii Novgorod and Novgorod region
- Bhutan Observer
- Bolivian Express Magazine
- Botswana Daily News(Asianet-Pakistan (Pvt) Ltd)
- Brunei News Gazette
- Bali Times (Indonesia)
- Business Telegraph
- Caloundra Weekly (Queensland)
- Calgary Sun
- Cambodia News Gazette
- The Capital (Annapolis, MD)
- Cape Times (South Africa)
- The Capital Times (Madison, Wisconsin)
- Central Queensland News
- Chicago Daily Herald
- China Daily - Africa Weekly
- China Daily (Hong Kong Edition)
- China Daily
- China Daily - US Edition
- The Columbian (Vancouver, WA)
- Crest
- Daily Dispatch (South Africa)
- Daily News(S. Africa)
- Daily Pakistan Today
- Daily Nation (Kenya)
- The Daily Record (Baltimore, MD)
- Daily National Herald Tribune
- Daily News (Sri Lanka)
- Daily News (South Africa)
- Daily Star
- Evening Chronicle (Newcastle,UK)
- Herald Sun/Sunday Herald Sun (Melbourne, Australia)

- Daily Independent (Nigeria)
- Business Times (Malaysia)
- Palestinian News Network (PNN) - English
- Qatar News Agency (QNA) (English)
- Daily Record and Sunday Mail
- Rossiyskaya Gazeta - Weekly
- The Sunday Star-Times (Auckland, New Zealand)
- The State Journal-Register (Springfield, IL)
- The Press (Christchurch, New Zealand)
- U.S. News & World Report
- Waikato Times (Hamilton, New Zealand)

## ***Appendix II: BARK and MAROB Coding Comparison***

**BARK Variable:**LEFT (Leftist Organization)

The organization promotes economically leftist policies such as redistribution of wealth by the government and nationalization of industry. Most that fall into this category are communist and socialist organizations and their variants (Marxist, Leninist, Maoist). Examples include the New People's Army (NPA), the Red Brigades and the Revolutionary Armed Forces of Columbia (FARC).

**MAROB Variable:**LEFTORG

Does the organization advocate policies to redistribute wealth (e.g., extensive welfare policies, state control of major industries)?

- 0 No
- 1 Yes (Leftist- economic)
- 99 Missing Value/No basis for judgment

**Comparison:** 370 MAROBID\_YEARs for 2005/2006

- 11 -99 values, coded as '0' in BARK
- Assumed blanks in MAROB\_EE were -99s

| MATCH # | PERCENTAGE |
|---------|------------|
| 322/370 | 87.03%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 319/344 | 92.73% |

**When coded '1' in MAROB:**

| MATCH # |     |
|---------|-----|
| 3/15    | 20% |

**BARK Variable:**

RGHT (Rightist Organization)

Organizations that promote economically rightist policies such as extreme laissez-faire policies, small government, or policies that favor status quo economic elites. Often rightist organizations are formed as a reaction to leftist organizations, such as many anti-communist organizations in Central and South America. Examples include United Self-Defense Forces of Columbia (AUC) and Mountaineer Militia.

**MAROB Variable:**

RIGHTORG

Does the organization advocate free market policies OR policies that favor traditional economic elites (e.g., landowners)?

0 No

1 Yes (Rightist - economic)

-99 Missing Value/No basis for judgment

**Comparison:** 370 MAROBID\_YEARs for 2005/2006

- 8/9 -99 values, coded as '0' in BARK
- Assumed blanks in MAROB\_EE were -99s

| MATCH # | PERCENTAGE |
|---------|------------|
| 325/370 | 87.84%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 321/341 | 94.13% |

**When coded '1' in MAROB:**

| MATCH # |     |
|---------|-----|
| 4/20    | 20% |

**BARK Variable:**

ETHN (Ethnic Organization)

Organizations that represent certain ethnic groups and advocate for the rights or expansion of rights of that ethnic group. Examples include Fatah, Basque Fatherland and Freedom (ETA), and the Kurdistan Workers' Party (PKK).

**MAROB Variable:**

ETHORG

Ethnicity – ethnic community but no claims to autonomy or independence

0 No

1 Yes

-99 Missing Value/No basis for judgment

**Comparison:** 370 MAROBID\_YEARS for 2005/2006

- 3 -99 values, coded as '0' in BARK
- Assumed blanks in MAROB\_EE were -99s

| MATCH # | PERCENTAGE |
|---------|------------|
| 188/370 | 50.81%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 176/195 | 90.26% |

**When coded '1' in MAROB:**

| MATCH # |       |
|---------|-------|
| 12/172  | 6.98% |

### **BARK Variable:**

#### **TERRCNTRL (Territorial Control)**

We define territorial control as existing when the organization is able to control movement into, out of, or within a given territory. In some instances organizations will perform functions or provide services, similar to that of a legitimate government. Territory may be controlled by threat or use of force, or if the government grants the organization the authority to do so. The territory must be a substantial area (city, region, etc.) and not just an organization occupying a building or a city block. Thus this excludes military bases and checkpoints. Though bases may indicate that the organization does have control over territory, they may also be covert and thus not exercising social control over civilians in the area of the base. Additionally, the organization must have control over the territory for more than a few days. For instance, seizure and loss of a town in the space of a week does not constitute territorial control.

0 Does not control territory

1 Controls territory

Inference rule: all -99s were coded as 0 in the data.

### **MAROB Variable:**

#### **ORGST9**

Administering rebel areas: Coded 1 if organization controls movement into/out of/through territory but does not establish governing structures or maintain infrastructure (e.g., roads). Coded 2 if organization sets up governing structures and/or maintains infrastructure.

0 Not used as a strategy

1 Minor/Infrequent strategy

2 Major/Frequent strategy

-99 Missing Value/No basis for judgment

**Comparison:** 370 MAROBID\_YEARS for 2005/2006

| <b>MATCH #</b> | <b>PERCENTAGE</b> |
|----------------|-------------------|
| <b>325/370</b> | <b>87.84%</b>     |

**When coded '0' in MAROB:**

| <b>MATCH #</b> |               |
|----------------|---------------|
| <b>322/361</b> | <b>89.20%</b> |

**When coded '1' or '2' in MAROB:**

| MATCH # |        |
|---------|--------|
| 3/9     | 33.33% |

**BARK Variable:**

SYM\_RESIS (Symbolic Resistance)

Was symbolic resistance used by the organization? This includes scattered acts of symbolic resistance (e.g. sit-ins, blockage of traffic, sabotage, symbolic destruction of property), political organizing activity on a substantial scale or mobilization for autonomy/secession by a minority-controlled regional government.

0 No

1 Yes

**MAROB Variable:**

ORGST1

Education & propaganda: Coded 1 if organization has monthly/quarterly radio or print publications; if they have 2 or fewer foreign offices; if they leaflet or otherwise distribute material less than weekly. Coded 2 if organization has daily/weekly radio or print publication; if they have 3 or more foreign offices; if they distribute materials on a weekly or more frequent basis; or if they satisfy two or more of the conditions to code 1

0 Not used as a strategy

1 Minor/Infrequent strategy

2 Major/Frequent strategy

-99 Missing Value/No basis for judgment

**Comparison:** 370 MAROBID\_YEARs for 2005/2006

- 9 -99 values, coded as '0' in BARK
- Assumed blanks in MAROB\_EE were -99s

| MATCH # | PERCENTAGE |
|---------|------------|
| 213/370 | 57.57%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 207/211 | 98.10% |

**When coded '1' or '2' in MAROB:**

| MATCH # |    |
|---------|----|
| 6/150   | 4% |

**BARK Variable:**

VIOL\_REPRESS (Violent State Repression)

Does the state engage in violent repression against the organization?

- 0 State is not using violent repression against the organization
- 1 State is using violent repression against the organization

**MAROB Variable:**

STATEVIOLENCE

Does the state use violence against the organization?

- 1 State is not using lethal violence against the organization
- 2 State is using periodic lethal violence against the organization
- 3 State is using consistent lethal violence against the organization

-99 Missing Value/No basis for judgment  
**Comparison:** 370 MAROBID\_YEARS for 2005/2006

**Comparison:** 370 MAROBID\_YEARS for 2005/2006

- 3 -99 values, coded as '0' in BARK
- There were no '1' values coded in MAROB\_EE
- Assumed blanks in MAROB\_EE were -99s
- Assumed phrase 'state is' meant a '2' or '3' should be coded

| MATCH # | PERCENTAGE |
|---------|------------|
| 34/370  | 9.19%      |

When coded '1' in MAROB:

| MATCH # |     |
|---------|-----|
| N/A     | N/A |

When coded '2' or '3' in MAROB:

| MATCH # |       |
|---------|-------|
| 34/367  | 9.26% |

**BARK Variable:**

RELI (Religious Organization)

Organizations guided by some form of religious principles. They may seek to incorporate religious policies into public life or exist to protect a distinct religious organization.

Examples include al-Qa'ida, Jemaah Islamiya, Kach, and the Lord's Resistance Army (LRA).

**MAROB Variable:**

RELORG

Does the organization advocate policies that incorporate religion into public life

0 No (non-religious)

1 Yes

-.99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if religious at any point in time period

| MATCH # | PERCENTAGE |
|---------|------------|
| 64/111  | 57.66%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 48/84   | 57.14% |

**When coded '1' in MAROB:**

| MATCH # |        |
|---------|--------|
| 16/27   | 59.26% |

**BARK Variable:**LEFT (Leftist Organization)

The organization promotes economically leftist policies such as redistribution of wealth by the government and nationalization of industry. Most that fall into this category are communist and socialist organizations and their variants (Marxist, Leninist, Maoist). Examples include the New People's Army (NPA), the Red Brigades and the Revolutionary Armed Forces of Columbia (FARC).

**MAROB Variable:**LEFTORG

Does the organization advocate policies to redistribute wealth (e.g., extensive welfare policies, state control of major industries)?

- 0 No
- 1 Yes (Leftist- economic)
- 99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if leftist at any point in time period

- one -99 value, coded as '1' in BARK

| MATCH # | PERCENTAGE |
|---------|------------|
| 72/111  | 64.86%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 60/83   | 72.29% |

**When coded '1' in MAROB:**

| MATCH # |        |
|---------|--------|
| 12/27   | 44.44% |

**BARK Variable:**

RGHT (Rightist Organization)

Organizations that promote economically rightist policies such as extreme laissez-faire policies, small government, or policies that favor status quo economic elites. Often rightist organizations are formed as a reaction to leftist organizations, such as many anti-communist organizations in Central and South America. Examples include United Self-Defense Forces of Columbia (AUC) and Mountaineer Militia.

**MAROB Variable:**

RIGHTORG

Does the organization advocate free market policies OR policies that favor traditional economic elites (e.g., landowners)?

0 No

1 Yes (Rightist - economic)

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if rightist any point in time period

| MATCH # | PERCENTAGE |
|---------|------------|
| 83/111  | 74.77%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 79/103  | 76.70% |

**When coded '1' in MAROB:**

| MATCH # |     |
|---------|-----|
| 4/8     | 50% |

**BARK Variable:**

NATILST (Nationalist)

Organizations that proclaim allegiance to a particular country and/or advocates national independence or a strong national government.

**MAROB Variable:**

NATORG

Nationalist—claims to autonomy or independence

0 No

1 Yes (Nationalist)

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if nationalist at any point in time period

- one -99 value, coded as '0' in BARK

| MATCH # | PERCENTAGE |
|---------|------------|
| 69/111  | 62.16%     |

**When coded '0' in MAROB:**

| MATCH # |        |
|---------|--------|
| 60/67   | 89.55% |

**When coded '1' in MAROB:**

| MATCH # |        |
|---------|--------|
| 9/44    | 20.45% |

**BARK Variable:**

ETHN (Ethnic Organization)

Organizations that represent certain ethnic groups and advocate for the rights or expansion of rights of that ethnic group. Examples include Fatah, Basque Fatherland and Freedom (ETA), and the Kurdistan Workers' Party (PKK).

**MAROB Variable:**

ETHORG

Ethnicity – ethnic community but no claims to autonomy or independence

0 No

1 Yes

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if ethnic at any point in time period

- one -99 value, coded as '0' in BARK

| MATCH # | PERCENTAGE |
|---------|------------|
| 47/111  | 42.32%     |

**When coded '0' in MAROB:**

| MATCH # | PERCENTAGE |
|---------|------------|
| 31/49   | 63.26%     |

**When coded '1' in MAROB:**

| MATCH # | PERCENTAGE |
|---------|------------|
| 16/62   | 25.81%     |

**BARK Variable:**

GENDEXC (Gender Exclusiveness)

Organizations who advocate the exclusion of women from public life.

**MAROB Variable:**

GENDEXC

Does organization advocate the exclusion of women from public life?

0 No

1 Yes (Gender exclusiveness)

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if practiced gender exclusiveness at any point in time period<sup>8</sup>

| MATCH # | PERCENTAGE |
|---------|------------|
| 106/111 | 95.49%     |

**When coded '0' in MAROB:**

| MATCH # |      |
|---------|------|
| 106/106 | 100% |

**When coded '1' in MAROB:**

| MATCH # |    |
|---------|----|
| 0/5     | 0% |

---

<sup>8</sup> BARK had no instances of gender exclusive organizations

**BARK Variable:**

SUPR (Supremacist Organization)

Organizations that advocate for the inherent supremacy of a specific race over other supposedly inferior races. Some extreme nationalist organizations reach this level as well. Examples include the Ku Klux Klan, Aryan Nations, and Russian National Unity.

**MAROB Variable:**

SUPREMACIST

Does organization advocate inherent superiority of racial/ethnic group and advocate for exclusive political power by that group?

0 No

1 Yes (Racially/ethnically supremacist)

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if supremacist at any point in time period<sup>9</sup>

| MATCH # | PERCENTAGE |
|---------|------------|
| 87/111  | 78.38%     |

**When coded '0' in MAROB:**

| MATCH # |      |
|---------|------|
| 87/111  | 100% |

---

<sup>9</sup> MAROB had no instances of supremacist organizations

**BARK Variable:**

ENVI (Environmental Organization)

Organizations that advocate for policies that favor the conservation of the environment or wildlife. Examples include the Earth Liberation Front (ELF), Animal Liberation Front (ALF), and Oxford Arson Group.

**MAROB Variable:**

ENVORG

Does the organization advocate environmentally sustainable policies?

0 No

1 Yes (Environmental)

-99 Missing Value/No basis for judgment

**Comparison:** 111 organizations, coded 1 if environmentalist at any point in time period<sup>10</sup>

| MATCH # | PERCENTAGE |
|---------|------------|
| 111/111 | 100%       |

---

<sup>10</sup> MAROB and BARK had no instances of environmental organizations

## Source Collection and Coding Challenges

### 1. Lexis-Nexis

- a. It took us a long time to get the agreements in place to allow us to get an account
- b. The Research Foundation was slow in paying the bills and Lexis-Nexis turned off our account until that was resolved. We lost a few weeks during that time.
- c. We were not properly controlling the limits and exceeded them on weekend. As a result, we lost access for 2 weeks, and had to come up with a quick fix. However, the bug allowed us to download about 6 million documents that weekend, which meant in the big picture we came out fairly even in the end.
- d. In a perfect world, we would be able to get 300K \* 4 (1.2M) document in a weekend. But because that is not the only limit on our processing, sources with long periods of no documents can eat up the searches and we may be throttled by the search limits and unable to maximize our download document counts, especially on Friday night and Monday morning. That is why we started running more than one search at a time. We may not get as far on a single source, but while one source might not be download documents, hopefully one of the others will.
- e. To help maximize our searches, we start each source at 20 days per search. But if the source starts getting errors about the search being too large, we cut the number of days back until we reach a number that works.
- f. Temporary network issues created gaps in the downloads. We had to create code to find, submit and reprocess the gaps. Some of the effort is manual at this time, but less than it was at the beginning of the project.
- g. Could have spent more time deciding what sources would have been the most useful and what order to process them in to maximize the amount of information we received from the data.
- h. The date range was changed from 2008-2018 to 2005-2018, to 2005-2017. Each change required some tweaking of the parameters and data. And for many of the sources, there was no additional data from 2005-2007, but to find that out, we used up our hourly search counts. We might have been able to determine details about the data by some other method, but we just did not have the resources.

### 2. Processing raw documents

- a. Creating the input files for the processor was time consuming, so a program was created to simplify the process. If we had to do it all over again, there are probably even more process improvements we could make.

### 3. Profiler

- a. The schemes were not available immediately. We had downloaded millions of documents before we processed a single one.
- b. There were quite a few errors in the schemes in some of the initial runs, which is usually expected. We restarted the whole thing a few times early on, but then were able to just address the errors for then on, and just resubmit the failures.

- c. To improve the efficiency, keep as many of the source documents on the same machine as the controller they will run from as possible. This will help the IO (and possibly the permissions).
  - d. Possible process improvement for the future might be to create something that controls monitoring and feeding PPBs to the controllers. You don't want too many in the directories at one time for efficiency, but you also don't want to run out. We did it manually.
- 4. Creating output
  - a. It took some time to get sample output.
  - b. We had some turnover on the part of the project involving the results between the time the output was specified and the time the output was used. This created some confusion on why the data looked like it did. This also generated some minor changes to the output format to help out in the analysis process.

### ***Appendix III Original Proposal***

i. Background and Objective

In the 1970s, an offshoot of the nonviolent Communist Party of Peru became the militant Shining Path organization, eventually responsible for countless acts of political violence in the past thirty years. In contrast, despite a violent history that includes an assassination attempt on President Nasser, the Muslim Brotherhood in Egypt renounced violence in the 1970s and took an increasingly nonviolent stance - until the recent military overthrow of their government. Why would a violent non-state organization decide to give up violence in favor of more democratic, less violent forms of political action? Conversely, why would a nonviolent organization decide to give up its nonviolent ideology in favor of political violence? An understanding of the factors that lead to these types of shifts in the use of violence would be an important part of governments' decision making process when deciding how to shape their policies towards these organizations.

Unfortunately, there is a key missing factor in much of our understanding of these questions. This missing factor is the lack of null cases – that is a lack of comparative analysis with groups that do NOT use violence or do not even use contentious tactics like protest. This makes it almost impossible to examine the cases where the *dog does not bark*. This problem is compounded by a general lack of robust data on organizational factors for nonviolent, non-contentious actors. Our research effort will correct this major research lacuna that is seriously hampering our understanding of the choice of political violence. We will build on the Minorities at Risk Organizational Database (MAROB) dataset, one of the only efforts that collected yearly data on both violent and nonviolent as well as non-contentious political organizations to better understand why organizations start being contentious, why they start using violence and why they stop. We will apply a variety of forecasting models and hazard models which have been rarely used in political science efforts on this topic. We will further innovate by working to automate the collection of organizational characteristics so that data can be collected continuously and paired with event data that is already being collected by other organizations.

ii. Scientific Approach and Relationship to competing or related research

Our proposed research program develops a theoretical framework for understanding the use of violence by non-state organizations and organizational decisions to escalate, de-escalate, or vacillate in their use of violence (Gurr, 1970; Rasler 1996). We focus on how organizational decisions are influenced both by inter-organizational dynamics and by government retaliatory and consolatory actions towards non-state groups within a polity. In particular, we argue that (a) pre-existing historical group ideological characteristics (religious, cultural, and political), (b) domestic and international political opportunity structures, and (c) intergroup coalition networks condition the initial decision of a non-state organization to use violence (Tilly, 1973; Lichbach, 1994; Gates, 2002 Berman and Laitin, 2008; Horowitz, 2010). Certain government retaliatory actions to domestic groups connected to the organization, like publically imprisoning nonviolent

activists, make a change from nonviolent to violent strategies more likely as these actions undermine trust in the government and credibility in government conciliatory gestures. Conversely, international organizational contacts and the government's expansion of association rights make a move towards nonviolent tactics more likely for non-state groups with a history of violence (Pierskalla, 2010, Siegel, 2011 Bell et al., 2012).

Using these theoretical priors, we will use a novel quantitative approach that allows us to simultaneously examine the ideology, behavior, and position of a non-state actor within both its organizational and organizational-country network. One of the key challenges when using quantitative methods to examine social and political stability is that there is very little data that captures non-state actor organizational factors (such as ideology, size, network connections, etc.) that may be related to movements or organizations that are challenging or supporting governments. To address this issue we will build on work done with one of the few existing datasets that captures the organizational data on both violent and nonviolent groups across the Middle East region – The Minorities at Risk Organizational Database (MAROB). These data allow us to identify key factors that help to explain why organizations (a) begin, continue, or end the use of violence and/or protest, (b) choose a mixed strategy that involves both violence and non-violence, and (c) the impact of government behavior. These factors are both contextual (state-based, culturally-based, regionally-based) and organizational (size, ideology, religion, leadership traits and the sentiments about the group in the society).

Previous MAROB research (as well as research using the related Political Organization Database that codes a sample of Middle Eastern Extremist organizations both violent and not) has used standard logistic analysis to analyze the decision to use violence as well as the decision to use different types of contentious political behavior from protest to violence (Asal et al. 2013, Asal et al. forthcoming1, Asal et al. forthcoming 2). This analysis has already shed new light on the choice of violence within the context of ethno-political organizations in the Middle East. For example, one analysis found that religious ideology is not impacting the decision to use violence as much as it is impacting the decision to target civilians once violence is chosen as a strategy (Asal et al. forthcoming 1). Another analysis underlined the importance that gender ideology can have on the impact of the choice of violence; in that organizations with such an ideology were much less likely to choose violence (Asal et al. 2013). Previous MAROB analysis also found that government behavior had a big impact on how organizations chose to behave in terms of the kinds of contentious tactics they chose (Asal et al. 2013, Asal et al. forthcoming1, Asal et al. forthcoming2). The fact that the MAROB data includes comparable violent and nonviolent organizations has also allowed for the ability to identify if a group's participation in violence has an impact on organizational splits (Asal et al. 2012a), as well as the propensity of an organization

to be involved in the illegal drug trade (Asal et al. 2012b), or the likelihood that they will engage in transnational activity (Asal et al. forthcoming 3).

However, despite these important results, the existing MAROB data and previous analyses of political violence have some important limitations that prevent definitive testing of many questions:

- Limited theoretical scope. Much of the existing research on terrorism focuses only groups that use violence and the violence they use (see, for example, Choi, Seung-Whan, 2010). This creates a problem for analysis in that null cases (groups that do not use violence) are missing and analysis may reveal effects that appear real but are in fact spurious. A full test of the causal importance of the identified factors requires the presence of a substantial number of null cases.
- Data Analysis. Most analysis of political violence within the extant literature employs some form of logistic analysis to model the forces that shape whether an organization is engaged in violence or not. Although such an approach provides useful information about factors connected to political violence, it only provides a narrow snapshot view of violence. In this snapshot, groups are either violent or they are not. A richer analysis would not only seek to understand the determinants of whether a group is violent or not, but also what influences *how long* the group continues to use violence as a strategy for achieving its goals.
- Limited data. The original MAROB data set runs from 1980 to 2004 and covers 118 ethno-political organizations in 16 countries of the Middle East and North Africa. To fully test theories that explain the use and non-use of violence the dataset should be extended in time to the present day and extended in scope to cover various types of political groups (including non-violent, non-contentious groups) around the entire globe. Unfortunately, as one unsuccessful attempt to extend MAROB demonstrates, replicating the original effort is extremely difficult, time consuming and costly.
- Missing data. The existing MAROB dataset does not contain network data beyond simple dyads, and our existing terrorist datasets only allow us to examine the relational dynamics of violent behavior. We have no window into the relational dynamics of non-violent behavior and transitional behavior (that is, the decision to switch from contentious to violent political means). For example, does rivalry drive organizations to choose violent behaviors, much as rivalry drives organizations already engaged in violence to engage in more extreme violence? Or does rivalry in some cases drive organizations to differentiate – that is, to choose nonviolent means so that the focal group serves as an alternative for those members of a minority group that wish to express dissatisfaction with government

action but not through violent means? Is violence “contagious”: if one organization representing an at-risk population chooses violence, does that draw all others into violence through outbidding? We know from analysis of the existing data that organizations transition back and forth between periods of violence and nonviolence; network relations may provide a missing explanatory piece. Fundamentally, we wish to examine the degree to which network and behaviors co-evolve across time using stochastic social network methods, but to pursue this line of analysis we need longitudinal data on alliances and rivalries among MAROB groups. MAROB currently treats each organization it covers in isolation as though it has no interaction with other groups.

This effort will address each of these limitations.

### **Theoretical Scope.**

Much of the previous research on terrorism has focused only groups that use violence and the violence they use (see, for example, Choi, Seung-Whan, 2010), our research however emphasizes that actors can choose multiple actions that are contingent upon organizational, ideological, and environmental factors. In order to have robust research design that includes null cases, we will identify groups that share ideological features and goals but that differ in that some use violence and some do not and thus identify those features that make violence more or less likely. This gives us greater leverage on the conditions under which political organizations tend to adopt violence and those that encourage them to abandon it.

In addition, we will combine theories of contentious politics and human rights with theories from social network analysis (SNA) to gain a better understanding of why organizations choose the strategies they do. Our approach draws on the literatures on grievance, resource mobilization, human rights, opportunity structure, and social networks (Gurr, 1970; Poe and Tate, 1994; Siegel, 2009, 2011). An organization’s relations with government and other competitive and allied organizations will interact with an organization’s capabilities, its leadership, its cultural context, and its ideology to determine likely courses of behavior. We will build on our previous findings: 1) that ideology (including gender ideology) and organizational size combined with network structure has a huge impact on an organization’s choice of violence or nonviolence and its lethality if it chooses violence (Asal and Rethemeyer, 2008a, 2008b; Asal, Ackerman, and Rethemeyer, 2012), and 2) that government policies and actions also hugely influence organizational choices (Asal, Brown, and Dalton, 2012; Dalton Asal, 2011).

### **Data Analysis**

Our data analysis team will move beyond the snapshot of logistic regression and pursue three related lines of inquiry: hazard analysis, network analysis, and co-evolution modeling. Team members J. Michael Greig and Joseph Young will use

hazard and survival analysis tools to analyze the determinants of organizational choice to use or eschew violence. Team members Amanda Murdie and R. Karl Rethemeyer will use network methodologies such as stochastic modeling that can identify how relationships are impacting behavior and how behavior impacts relationships in the future. All the data analysis team members will collaborate on co-evolution modeling.

Hazard Analysis. In contrast to the single snapshot provided by logistic regression, hazard analysis allows us to explore the factors that influence the duration that organizations avoid the usage of violence as well as those factors that influence the initial adoption of violence. Fundamental to hazard analysis is its ability to deal with “censored” observations. Hazard analysis models the time until an event occurs for an actor, such as the adoption of violence by political organizations. Some actors, however, never see such an event during the time they are under study (i.e.; never become violent). Hazard analysis provides a means of explicitly modeling these censored cases, avoiding the need to drop them from the analysis and avoiding the selection bias that such an exclusion would introduce into the analysis.

At the same time, the hazard analysis approach also allows us to model the factors that shape how long groups that have adopted violence remain violent. In this respect, the hazard analysis approach will allow us to not only model and predict which groups become violent, but also when they become violent and how long they are likely to continue to engage in violence. This provides powerful empirical leverage in understanding the forces that influence both the initiation and termination of violence by groups. This leverage, in turn, can allow us to develop specific forecasts for the adoption and continued use of violence by individual ethnopolitical organizations. Factors that shorten the duration until the onset of violence are those that increase the propensity of groups to turn toward violence while those that lengthen the duration until the adoption of violence are those that undermine the tendency of groups to become violent.

Adopting a hazard analysis approach brings three key benefits with it. First, it allows us to analyze over time all groups that may potentially become violent, not just those that actually become violent. This allows for our analysis to avoid problems of selection bias in modeling the occurrence of violence among groups. Second, it gives us greater empirical leverage over the strength of influence that individual factors have upon the propensity of groups to adopt violence. Third, the use of hazard analysis will allow us to also analyze the duration that groups continue the use of violence once it is adopted. Because groups that become violent can sometimes transition away from violence, our approach will allow us to also analyze how factors shape the duration that groups continue to use violence once they adopt it in the first place.

For our hazard analysis of violence by political organizations, we will need network data to create organization-level indicators of network structure and position – for instance, centrality measures, coreness measures, measures of ego-network size, etc. This research will focus on factors that help to explain whether an organization is likely to ever select violence, the factors that expand or contract “spells” of violence or non-violence, factors that explain the first use of violence, factors that promote demobilization once violence occurs, and factors that help to explain the intensity of violence once violence is selected. The models will use context-specific variables, organizational variables, and network variables. Modeling will also account for unit and time effects generated by the use of panel data.

Network Analysis. Our work on terrorist organizations (Asal and Rethemeyer 2008, Asal, Ackerman, and Rethemeyer 2012) has demonstrated repeatedly that positive network connections help to explain organizational behavior (such as lethality and pursuit/use of CBRN capabilities). We also know from more recent work with longitudinal network data on terrorist organizations that negative relationships – rivalry, rhetorical hostility, intergroup armed conflict – leads to intensification of these behaviors (Asal, Phillips, and Rethemeyer, in review) through processes that echo the work on terrorist outbidding (Bloom 2005). We also know that violent behaviors help to influence the nature and array of network relationships terrorist groups form (Asal, Park, Rethemeyer, and Ackerman, forthcoming) and thus that there is an important causal loop: network relationships beget violence and violence begets relationships. This research will focus on the nature of organizational and organizational-country networks and will seek to understand patterns of cooperation and conflict among political organizations and between organizations and countries. Special attention will be paid to both domestic and international organizational contacts. Modeling here will draw on recent work (Ripley, Snijders and Preciado 2011; Snijders 2001; Snijders 2002; Snijders 2005; Snijders and Duijn 2002) on longitudinal social network modeling to discover factors that explain entry, exit, tie formation, and tie dissolution over time. In this case the network is the dependent variable with organizational choices (use of protests, use of violence, choice to end violence, etc.) and organizational characteristics, and contextual factors as the independent variables. By adding network coding to an analysis of violent and nonviolent coding we will be able to see how network relations impact the choice of violence. This project will close a fundamental gap in our current understanding of contentious political groups: the extent to which the network of relationships between such organizations influences their behavior.

Co-evolution. In this analysis, the focus will be on the simultaneous determination – or “co-evolution” – of networks and behavior, with behavior driving network “selection” and network structure influencing behavior. The preceding work on hazard analysis and social networks will greatly increase the chances of success in this line of inquiry, as co-evolution models are difficult to properly specify and estimate. We will focus on both the behavioral variables outlined in the first line of

inquiry and the network structure variables outlined in the second line of inquiry. This analysis will allow us to better understand the ways in which the networks that organizations embed themselves shape their propensity to adopt violence and, once they do so, their likelihood of stopping violence. Co-evolution modeling is new, and the size and complexity of the datasets that this project will generate will stretch the capabilities of the software that is available for this line of inquiry. Thus success in this final line of inquiry is a stretch goal for the project, with the first two lines of inquiry being the primary approaches.

### **Data scope and resolution.**

The data analysis efforts we envision require data with a broader scope and finer temporal resolution than provided by any existing dataset, including MAROB. In particular our analysis will require two related and coordinated datasets: (1) a case-by-variables dataset that includes characteristics of the organizations and (2) a new network dataset that includes information on both supportive and conflictual relationships between and among, (a) the organizations and (b) the organizations and countries in the case-by-variables dataset. Furthermore, to make the best use of the data and the proposed analysis techniques data at the temporal resolution of the month will be significantly more useful than data at the temporal resolution of the year.

However, the cost of producing the necessary dataset by traditional human coding of source documents would be very expensive and very slow. For example a comparable effort that is generating updates to the Big Allied and Dangerous dataset will generate data at a cost of nearly \$700,000. This type of data collection effort is increasingly difficult to sustain and justify in an environment of limited research budgets and large numbers of worthy projects competing for those funds. In order to address these issues and to allow the project to continue into the future we will create a generalizable methodological approach that reliably and accurately measures the hypothesized causal factors by automating the coding of those factors and thereby greatly extend the scope, accuracy, and reliability of data collection while simultaneously dramatically reducing the cost.

Automating text coding has several distinct advantages over manual methods of data collection including:

- Reliability.
- Speed.
- Flexibility.
- Scope.
- Re-usability.
- Cost.

These advantages have led to a commercial market for automated text coding that is approaching \$1 billion a year (Grimes, 2011).

### Reliability

As Neuendorf (2002) notes, “given that a goal of content analysis is to identify and record relatively objective (or at least intersubjective) characteristics of messages, reliability is paramount. Without the establishment of reliability, content analysis measures are useless.” “Coefficients of .90 or greater would be acceptable to all, .80 or greater would be acceptable in most situations, and below that there exists great disagreement”. Despite the importance of coding reliability for the interpretation of research results, only 69% of research reports using content analysis report any intercoder reliability (Lombard, Snyder-Duch and Bracken, 2002) and the levels reported in *published* articles can be as low as .40. Fortunately automated systems have a built-in reliability of 1.0.

### Speed

Profiler Plus achieves excellent throughput and has a scalable architecture sufficient to handle enormous volumes of text in near real-time. Actual throughput achieved by Profiler Plus depends on: the hardware, the coding scheme used and the input text file size. On a single 2.8 GHz Pentium processor with 1 GB of RAM, and using an English language sentiment analysis coding scheme, throughput ranges from 2 to 5 MB (roughly 1,000 to 2,500 pages) per hour for a sample of 148,133 articles (4.6 GB) from the New York Times, the Washington Post, The Wall Street Journal, and USA Today (print and on-line). As King and Lowe (2003) observe, an automated coding system can “code millions of news stories in the time it takes a human being to code only a few” (King and Lowe 2003).

### Scope

A direct consequence of the speed of automated coding is that it allows researchers to significantly broaden the scope of their research and, at the same time, increase the resolution. Automated systems are capable of pulling large amounts of data from the nearly endless stream of text made available on the internet every day or from widely available large news archives. For example, the BBC Worldwide Monitoring online repository collects documents from a large variety of source news outlets spanning the entire globe, generally pulling over twenty thousand documents per month. Other sources, including chat forums, blogs, and other news outlets, can be automatically downloaded regularly and mined for information.

### Flexibility

A frequent occurrence in text coding projects is the discovery of a new term or phrase that should be coded but has previously escaped notice. This discovery can be agonizing if made at the end of a human coded project as the discovery requires a new rule to be added to the coding scheme and for all previously coded text to be reviewed for the new term or phrase. A daunting prospect when megabytes or gigabytes of text are involved. For automated systems the issue is simply a matter of encoding a new rule and rerunning the text. Similarly, if a new concept or

variable is introduced to the analysis the text can be re-run and extended data created with very little cost in a brief period of time.

### Re-usability

Turnover in human coders is usually unavoidable: students graduate, researchers move on to other institutions or responsibilities, and new coders must therefore be recruited and trained in a time-consuming and expensive process. This is not true for automated systems. Automated systems can be kept indefinitely in storage and used at any point in the future, although minor changes may be needed to adapt them to more recent operating systems.

### Cost

In the short term and for small volumes of text, automated coding systems can be more expensive than human coders, because we can take advantage of the deeply ingrained but implicit knowledge of language that most human beings have to code small volumes quickly with relatively little training—most of the necessary training is already present. However, as the volume of material increases the cost advantage shifts towards computers that can code 24 hours a, 7 days a week for pennies an hour and away from human coders that cost at least \$7.25 an hour and who code a tiny fraction of the text that an automated system codes in the same period. Although new automated coding schemes are expensive to develop and validate, once they have been developed, they are inexpensive to maintain and to update. In addition, once an automated coding scheme has been designed and tested, it can be applied as many times as one wishes with little change needed to the infrastructure.

To date, the use of automated coding in Political Science research has focused on generating event data. In this project, we will extend the scope of automated coding and collect both organizational features of the organization like ideology and size, as well as network relations of the organizations. In addition, the use of automated coding will allow us to generate data at a temporal resolution of the month rather than the year and to update and extend the dataset on a regular basis at very low cost.

#### iii. Level of effort

Principal Investigator Victor Asal's level of effort varies each year during the three year project. During the summer of the first year he is budgeted at .083 months or 2.77% effort ( $.083/3 = 2.77\%$ ). During the summer of the second year he is budgeted at .081 months or 2.70% ( $.081/3 = 2.70\%$ ). Finally, during the summer of the third year he is budgeted at .399 months or 13.30% ( $.399/3 = 13.30\%$ ). Dr. Asal will oversee the project, conduct analysis, and oversee the graduate and undergraduate students throughout the project.

Co-Principal Investigator Karl Rethemeyer is budgeted at the equivalent of .302 months or 10.06% effort during the summer of the third year ( $.302/3 = 10.06\%$ ). Dr.

Rethemeyer will assist with overseeing the project and conduct analysis, including social network analysis, in the final year of the project.

The graduate student will be working at 50% effort (20 hours per week) for 52 weeks per year for the first two years of the project. The graduate student will conduct coding and analysis and assist with overseeing the undergraduate coders.

Shared undergraduate hours are budgeted at 2,800 hours in the first year, 1,975 hours in the second year, and 2,000 hours in the third year. This is equal to 134.62% effort in the first year ( $2,800/40/52 = 134.52\%$ ), 94.95% in the second year ( $1,975/40/52 = 94.95\%$ ), and 96.15% in the third year ( $2,000/40/52 = 96.15\%$ ). These hours will be shared among multiple undergraduates each year. Each individual undergraduate student is likely to be working no more than 50% effort (20 hours per week).

iv. Nature and extent of anticipated results and Contribution to mission of Army

The results of this research will be in direct agreement with the mission of the Army's Social and Behavior Science program. The more granular and extensive data that we will collect, along with the spatio-temporal predictive modeling, will provide several useful tools for the DoD. In particular, the refined data will provide greater insight into sub-regions of countries in which ethnic and extremist organizations will have incentives to press for initial change, and also learn from the changes of others.

As a result, this research will allow us to answer the question about what factors lead organizations to switch between nonviolent and violent behavior, or vice versa. This will provide the DoD, and the U.S. Government in general, with useful information when deciding on how to interact with such groups in the future and help predict when a nonviolent group may be heading down the path to violence.

v. Identify any scientific uncertainties and describe approaches for addressing them.

There are two known scientific uncertainties in this research: 1) the applicability of the analytic methods and 2) our ability to develop a robust, accurate and reliable automated coding system for the data we seek.

Applicability of methods. Previous research in this area has used standard logistic analysis. We hypothesize that more recent techniques, such as hazard analysis, survival modeling, and social network analysis and modeling will provide more useful predictions. We may be wrong. These techniques have not been extensively applied to the types of data or contexts we are exploring. Co-evolution modeling in particular, is new, and the size and complexity of the datasets that this project will generate will stretch the capabilities of the software that is available for this line of inquiry. Thus success in this co-evolution modeling is a stretch goal for the project. In order to mitigate the risk from applying these more recent techniques we will also use the better known and previously used techniques such as logistic regression

to provide the greatest opportunity to identify causal relationships and generate useful predictions and policy guidance.

Automated Coding of Data. Although automated coding of event data is fairly widespread and growing due to the cost and speed advantages it offers over manual coding (see for example the W-ICEWS project), many in the Intelligence Community are not ready to use automated coding because of high levels of noise in the data (GDELT in particular suffers from this problem. See <http://johnbeiler.org/blog/2013/10/28/noise-and-gdelt/>, also personal communication with Jason Matheny at IARPA). It is reasonable to assume that automated coding of MAROB data may also suffer from noise in the data and potentially problematic accuracy. In addition, the existing MAROB coding scheme presents some other challenges for the development of an automated coding scheme.

Automated coding scheme development projects rely on the availability of some or all of the following:

1. Expert human coders
2. Detailed coding manuals
3. Substantial volumes of annotated source documents (“gold standard” data)
4. Documented relationships between a large number of original source documents and specific coded data.

Unfortunately, only a single human expert coder (Asal) is available to assist in the development of an automated MAROB coding scheme. In addition no annotated source documents currently exist, and the source document identifiers for specific coded data were not incorporated into the data set.

To overcome these obstacles, we will:

- 1) Document everything. To ensure replication we will extensively document the relationship between source documents and coded data and embed that information in the data set itself.
- 2) Create “gold standard” data. Because of the absence of coded data annotated with source identifiers or source data annotated with codes, Victor Asal, one of the original MAROB project directors will train a new cohort of MAROB hand coders at Albany and use those coders to generate a substantial number of MAROB coded documents using an in-line annotation format. The new MAROB data will retain the same variable codes as previous MAROB datasets but the exact location of the words in the source text that justify the codes will be identified explicitly—the variable codes will immediately precede the words or phrase that underlie the coding. These annotated documents will be used for automated testing of the new automated coding scheme. Half of the resulting documents will be used to develop the automated coding scheme and half will be “gold standard” data used to provide out of sample assessment of the coding scheme’s accuracy. As this

process occurs, we will also develop a new coding manual with extensive examples of correct and incorrect codes along with explanations. The result of this effort will be an improved codebook, a new coding manual and data which will provide a solid foundation for automation using Social Science Automation's *Profiler Plus* and associated utility programs.

- 3) Limit the scope of development. In order to ensure that robust coding processes and support materials are developed we will limit ourselves to developing guidelines and coding for 5 of the MAROB variables in the first year and to an additional 5 to 10 in the second year. Although this approach will only provide coding for 20% of the variables in the original MAROB dataset, we will select variables that have previously been found to have an impact on the use of violence as a way of verifying and validating the new automatically coded data.
- 4) Validation and Testing. Because we do not know the original source documents that gave rise to the existing MAROB data set, we cannot directly compare, and thereby validate, newly generated data against the existing data. In addition, we do not know whether the BBC Monitoring archive we will use for this project has generally the same information density and coverage as the range of sources used for the MAROB project. Fortunately, the BBC Monitoring archive we have access to and the MAROB data set overlap for the years 1997 to 2004, providing 8 years of comparison data. This overlap will allow us to assess the performance of the coding scheme and the information density of the BBC Monitoring archive. One major risk is that the BBC Monitoring service is not a rich enough source of data to replicate MAROB. Because of this risk, we will not expend the funds to acquire source documents from BBC Monitoring for Aug 2010 to present until we are confident that it provides sufficient coverage and information density.

vi. Description of organization

The University at Albany, State University of New York (SUNY) is the primary administrative contact for this project. However, all external grants are awarded to and managed by the Research Foundation (RF) for SUNY. The RF for SUNY is a private, nonprofit educational corporation that works closely with the SUNY campuses to manage and promote research, education, and discovery in the SUNY system. Vice Presidents for Research oversee all aspects of research on their campuses and serve as a liaison between the campuses and the RF. The Vice President for Research at the University at Albany is James Dias.

The Rockefeller College will provide the facilities needed to conduct this research project, including offices and work space for the PI Asal, computer access and computer networking, and access to relevant library and database services.

Rockefeller College has several research centers and institutes that conduct state-of-the-art research and policy analysis. The Center for Policy Research is primarily located in the main building of the Rockefeller College (Milne Hall) and has a

secondary location in a connected building (Draper Hall) sufficient to house all of the student workforce. PI Asal and professional support staff are currently located in Milne Hall.

- vii. Names of federal, state, local agencies, or others receiving this proposal and/or funding the proposed effort

A similar project with a focus on Africa and a wider array of groups, rather than the Middle East, will be proposed to the Minerva Research Initiative (Office of Naval Research).

- viii. Statement of impact on environment considering as a minimum water, atmosphere, natural resources, human resources, and any other values

Our project will not result in any impact on the environment (water, atmosphere, natural resources, human resources, or any other values).

- ix. Statement regarding the use of Class I and II ozone depleting substances

There are no Class I or Class II ozone depleting substances or research involving Class I or Class II ozone depleting substances.

- x. The type of support, if any, requested (e.g. facilities, equipment, and materials).

As reflected in the budget, the support requested in this proposal includes money for: salary and fringe benefits for graduate and undergraduate students at the University at Albany, travel for University of Albany personnel to Hilliard, Ohio and Washington, DC, subject matter expertise support, and for the sub award, Social Science Automation (SSA).