Mershon Center Data in Contentious Politics Workshop

18-19 January 2022

Participants

Note: participant and project descriptions are summarized from linked websites, journal articles, and reports. Some descriptions are abridged for brevity.

Jessica Maves Braithwaite

Dr. Braithwaite is an Associate Professor of Political Science at the University of Arizona. Braithwaite studies the dynamics of civil war and non-violent resistance, including organization mobilization in contexts of state repression and domestic unrest, as well as civilian experiences during conflict.

Website

Key topics: rebel groups; social organizations; human-coded data.

Selected Projects:

Foundations of Rebel Group Emergence (FORGE, with Kathleen Gallagher Cunningham)

The Foundations of Rebel Group Emergence (FORGE) Dataset examines the roots of rebellion by considering the characteristics and activities of the "parent" organizations from which rebel groups emerged, as well as some details about these rebel groups at the onset of conflict (the organization's "birthdate" and founding location, initial goals, ideology, and ethnic/religious foundations).

Details and Data
Introductory Article

Anatomy of Resistance Campaigns (ARC, with Charles Butcher)

This project focuses on the connections between and organizational/compositional characteristics of actors involved in maximalist anti-government campaigns as well as the formal social organizations (political parties and movements, trade unions, religious groups, student organizations, etc.) that mobilize alongside those rebel groups and comprise nonviolent movements.

Details and Data
Introductory Article

Erica Chenoweth

Dr. Chenoweth is the Frank Stanton Professor of the First Amendment at Harvard Kennedy School and a Susan S. and Kenneth L. Wallach Professor at the Radcliffe Institute for Advanced Study. Chenoweth studies political violence, non-violent resistence, and terrorism and directs the Nonviolent Action Lab at Harvard's Carr Center for Human Rights Policy.

Website

Key topics: non-violent resistance; campaign-level data; event-level data; human-coded data.

Selected Projects:

Nonviolent and Violent Campaigns and Outcomes (NAVCO)

The Nonviolent and Violent Campaigns and Outcomes (NAVCO) data project is the first of its kind to collect systematic data on both violent insurgencies and nonviolent civil resistance campaigns. The coverage is global but is limited to maximalist campaigns (i.e. those seeking to overthrow an incumbent government, expel a foreign military occupation, or secede). There have been several series of NAVCO, including:

Details and Data
Introductory Article (NAVCO 2.0)
Introductory Article (NAVCO 3.0)

- NAVCO 1._, which currently covers 622 campgains from 1900-2019.

- NAVCO 2._, which disaggregates campaigns to the yearly level.

- NAVCO 3._, which includes daily resistence events in 26 countries from 1990-2011.

Women in Resistance (WiRe, with Zoe Marks)

Details and Data
Associated Report

The Women in Resistance (WiRe) Dataset catalogues women's participation in 338 maximalist resistance campaigns (i.e. those campaigns that call for the toppling of an oppressive government or territorial self-determination). The dataset identifies both nonviolent and violent maximalist campaigns in every country in the world from 1945-2014, providing a comprehensive and systematic look at various dimensions of women's participation in both types of campaigns. (2019-04-29).

Crowd Counting Consortium (CCC, with Jeremy Pressman)

Details and Data
Associated Reports

The Crowd Counting Consortium (CCC) collects publicly available data on political crowds reported in the United States, including marches, protests, strikes, demonstrations, riots, and other actions.

Killian Clarke

Dr. Clarke is an Assistant Professor in the Edmund A. Walsh School of Foreign Service at Georgetown University, affiliated with the Center for Contemporary Arab Studies. Clarke studies the origins and consequences of grassroots mobilization and protests, and their contribution to transformative political events like revolutions, regime change, and democratization.

Website

Key topics: protest; revolution; local-sourced data; social media.

Selected Projects:

Which Protests Count? Coverage Bias in Middle East Event Datasets. Mediterranean Politics, 2021.

Article

Since the 2011 Arab Spring revolutions many scholars of the Middle East have built and analyzed locally-sourced protest event datasets, which have been hailed for providing superior coverage to various off-the-shelf datasets that rely primarily on English-language sources. This paper assesses the extent of these coverage improvements. It shows that across five different MENA countries, locally-sourced datasets identify considerably more events than most off-the-shelf datasets. It then compares one locally-sourced dataset of protests in Egypt from January 2012 to July 2013 to two prominent off-the-shelf datasets: ACLED and SCAD. These comparisons reveal that both ACLED and SCAD significantly overcount large, urban, violent, and political events. Next the paper compares the Egypt dataset to data compiled by two Egyptian activist groups, and finds that the locally-sourced dataset is also biased in key respects, undercounting small labor events outside the capital. Finally, the paper demonstrates the implications of these biases by

showing how statistical models of protest repression differ when using the locally-sourced dataset versus SCAD. Scholars of Mediterranean politics analyzing within-case and sub-national mobilization dynamics should use locally-sourced datasets whenever possible, but should also be aware that using local sources does not entirely eliminate certain forms of bias.

Launching Revolution: Social Media and the Egyptian Uprising's First Movers. BJPS, 2021.

Article

Drawing on evidence from the 2011 Egyptian uprising, this article demonstrates how the use of two social media platforms – Facebook and Twitter – contributed to a discrete mobilizational outcome: the staging of a successful first protest in a revolutionary cascade, referred to here as 'first-mover mobilization'. Specifically, it argues that these two platforms facilitated the staging of a large, nationwide and seemingly leaderless protest on 25 January 2011, which signaled to hesitant but sympathetic Egyptians that a revolution might be in the making. It draws on qualitative and quantitative evidence, including interviews, social media data and surveys, to analyze three mechanisms that linked these platforms to the success of the January 25 protest: (1) protester recruitment, (2) protest planning and coordination, and (3) live updating about protest logistics. The article not only contributes to debates about the role of the Internet in the Arab Spring and other recent waves of mobilization, but also demonstrates how scholarship on the Internet in politics might move toward making more discrete, empirically grounded causal claims.

David Cunningham

Dr. Cunningham is a Professor in the Department of Government and Politics at the University of Maryland. Cunningham's research focuses on civil war, conflict bargaining, international conflict management, conflict prevention, and international security.

Website

Key topics: civil war; non-state actors; data integration.

Selected Projects:

Non-State Actor Data (with Kristian Skrede Gleditsch and Idean Salehyan)

Details and Data
Associated Article

The Non-State Actor Data expand the information in the Uppsala Armed Conflict Data about civil conflicts in two directions: 1) information about the non-state actors involved in civil conflict and 2) information about the external dimensions conflicts, including external support and extraterritorial features.

Matching Event Data by Location, Time, and Type (MELTT, with David Backer, Karsten Donnay, Eric Dunford, and Erin McGrath)

Details
Associated Article

The growing multitude of sophisticated event-level data collection enables novel analyses of conflict. Even when multiple event data sets are available, researchers

tend to rely on only one. We instead advocate integrating information from multiple event data sets. The advantages include facilitating analysis of relationships between different types of conflict, providing more comprehensive empirical measurement, and evaluating the relative coverage and quality of data sets. Existing integration efforts have been performed manually, with significant limitations. Therefore, we introduce Matching Event Data by Location, Time and Type (MELTT)—an automated, transparent, reproducible methodology for integrating event data sets. For the cases of Nigeria 2011, South Sudan 2015, and Libya 2014, we show that using MELTT to integrate data from four leading conflict event data sets (Uppsala Conflict Data Project–Georeferenced Event Data, Armed Conflict Location and Event Data, Social Conflict Analysis Database, and Global Terrorism Database) provides a more complete picture of conflict. We also apply multiple systems estimation to show that each of these data sets has substantial missingness in coverage.

Cassy Dorff

Dr. Dorff is an Assistant Professor in the Political Science Department at Vanderbilt University and affiliated faculty at the Data Science Institute. Dorff conducts research on political conflict, resistance movements, collective action, networks, and political methodology.

Website

Key topics: political conflict; networks; analysis of machine-collected data; machine learning.

Selected Projects:

[W]hat Lies Beneath: Using Latent Networks to Improve Spatial Predictions (with Max Gallop and Shahryar Minhas)

Article

Spatial interdependencies commonly drive the spread of violence in civil conflict. To address such interdependence, scholars often use spatial lags to model the diffusion of violence, but this requires an explicit operationalization of the connectivity matrices that represent the spread of conflict. Unfortunately, in many cases, there are multiple competing processes that facilitate the spread of violence making it difficult to identify the true data-generating process. We show how a network-driven methodology can allow us to account for the spread of violence, even in the cases where we cannot directly measure the factors that drive diffusion. To do so, we estimate a latent connectivity matrix that captures a variety of possible diffusion patterns. We use this procedure to study intrastate conflict in eight conflict-prone countries and show how our framework enables substantially better predictive performance than canonical spatial-lag measures. We also investigate the circumstances under which canonical spatial lags suffice and those under which a latent network approach is beneficial.

Antigovernment Networks in Civil Conflicts: How Network Structures Affect Conflictual Behavior (with Nils W Metternich, Max Gallop, Simon Weschle, and

Michael D Ward)

Article

In this article, we combine a game-theoretic treatment of public goods provision in networks with a statistical network analysis to show that fragmented opposition network structures lead to an increase in conflictual actions. Current literature concentrates on the dyadic relationship between the government and potential challengers. We shift the focus toward exploring how network structures affect the strategic behavior of political actors. We derive and examine testable hypotheses and use latent space analysis to infer actors' positions vis-à-vis each other in the network. Network structure is examined and used to test our hypotheses with data on conflicts in Thailand from 2001 to 2010. We show the influential role of network structure in generating conflictual behavior.

Comparing GDELT and ICEWS event data (with Michael D Ward, Andreas Beger, Josh Cutler, Matthew Dickenson, and Ben Radford)

Report

This report offers an initial comparison of two major automated event data collection efforts – GDELT and ICEWS. Among other findings, the results indicate that GDELT tends to overstate contentious event counts, while ICEWS likely misses some events.

Laura Dugan

Website

Dr. Dugan is Ralph D. Mershon Professor of Human Security and Professor of Sociology at The Ohio State University. Dugan's research is motivated by the broader question of how leaders can reduce or enhance the risk of violence and other types of insecurities due to extremist ideologies and hateful intent.

Key topics: terrorism; open-sourced data; automated + human coded-data.

Selected Projects:

Government Actions in Terror Environments (GATE, with Erica Chenoweth)

Details and Data (GATE-Israel)
Introductory Article
Associated Article

Much effort has been exerted to develop terrorist incident databases that capture details of terrorist attacks across the globe. With these data, scholars and policy experts can observe patterns of attacks across regions, by specific terrorist organizations, and over time. Further, such data allow us to assess the relationship between possible causal factors, such as political climate and economy on the inception and rise of terrorist violence. These databases also allow us to assess the effectiveness of government interventions on reducing terrorism. Consequently, quantitative analysis of terrorism has grown substantially over recent years. Yet, when we raise the question of what does and does not work to reduce terrorist violence, we are limited to only those interventions that are explicitly publicized as counterterrorism. Missing from analysis are government actions that fall outside the purview of counterterrorism, yet plausibly affect terrorist violence either directly through the organizations or indirectly through their constituencies. This

paper introduces a new way to collect data on what governments do and presents descriptive accounts of conciliatory and repressive actions by governments relative to terrorist attacks in several countries. We also present an overview of findings that assesses the effects of conciliatory and repressive actions that are targeted both discriminately and indiscriminately on terrorist violence.

Global Terrorism Database (with Garry Lafree)

Details and Data
Introductory Article

The Global Terrorism Database (GTD) is an open-source database including information on domestic and international terrorist attacks around the world from 1970 through 2019, and now includes more than 200,000 cases. For each event, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and–when identifiable–the group or individual responsible.

David Muchlinski

Dr. Muchlinski is an Assistant Professor of international affairs in the Sam Nunn School of International Affairs at Georgia Tech. Muchlinski's research agenda is centered around the determinants of human pro-sociality, or why humans decide to behave cooperatively or violently with one another. Muchlinski's current research examines the correlates of extreme forms of political violence including genocide and other mass atrocities.

Website

Key topics: electoral violence; mass atrocities; machine learning; social media.

Selected Projects:

We need to go deeper: measuring electoral violence using convolutional neural networks and social media. PSRM, 2021 (with Xiao Yang, Sarah Birch, Craig Macdonald, and Iadh Ounis).

Article

Electoral violence is conceived of as violence that occurs contemporaneously with elections, and as violence that would not have occurred in the absence of an election. While measuring the temporal aspect of this phenomenon is straightforward, measuring whether occurrences of violence are truly related to elections is more difficult. Using machine learning, we measure electoral violence across three elections using disaggregated reporting in social media. We demonstrate that our methodology is more than 30 percent more accurate in measuring electoral violence than previously utilized models. Additionally, we show that our measures of electoral violence conform to theoretical expectations of this conflict more so than those that exist in event datasets commonly utilized to measure electoral violence including ACLED, ICEWS, and SCAD. Finally, we demonstrate the validity of our data by developing a qualitative coding ontology.

Countries at Risk of Electoral Violence (CREV, with Sarah Birch)

Details and Data
Introductory Article

The dataset of Countries at Risk of Electoral Violence (CREV) provides detailed dyadic information on electoral violence in 101 countries between 1995 and 2013. For an election to be deemed "at risk" of electoral violence, two criteria have to be met. The country in which the election has taken place must not have been a fully consolidated democracy (defined as having a Polity IV (Marshall, Gurr and Jaggers 2016) score of 10) throughout the entire time period covered by the data, and it must have sufficient media coverage (defined as an average of at least 365 reported events per year in the ICEWS dataset (see below for details)). The dataset of Countries at Risk of Electoral Violence follows the National Elections across Democracy and Autocracy (NELDA) election classification (Hyde and Marinov 2012; 2014). Elections in CREV are for national-level legislative and executive contests only, local and regional elections are excluded, as are referendums and constituent assembly elections. Electoral violence is measured in a ten-month window around each election. We code violence beginning six months before the election, three months after the election, and the month of the election. We provide two versions of the dataset. One is a time series cross-sectional (TSCS) dataset in which the unit of observation is the election, and where events of electoral violence are summed during the ten-month window. The other is a time series cross-sectional (TSCS) dataset in which the unit of observation is the electoral cycle month, and counts of violent events are specific to a given month during an electoral cycle.

Javier Osorio

Dr. Osorio is an Assistant Professor in the School of Government and Public Policy at the University of Arizona. Osorio's research interests focus on understanding the micro-foundations and dynamics of political and criminal violence in Latin America. Osorio uses several quantitative tools, including natural language processing, quasi-experimental and experimental techniques, geographic information systems, and big data analytics.

Website

Key topics: political and criminal violence; non-English sources; natural language processing; automated event coding.

Selected Projects:

Organized Crime Violence Event Data in Mexico (OCVED)

Details and Data
Associated Article

The Organized Crime Violence Event Data in Mexico (OCVED) is a large database containing events of drug-related violence in Mexico at the municipal level from 2000 to 2010. The data is machine-coded using `Eventus ID`, which adapts the `Tabari` algorithm for Spanish language sources. The data draws on 105 sources, including federal and local government agencies and national and local newspapers.

Translating CAMEO verbs for automated coding of event data

Associated Article
`Eventus ID` (for Spanish)
`Hadath` (for Arabic)

The Conflict and Mediation Event Observations (CAMEO) framework is the industry standard for computerized event coding that produces massive data on

political conflict at a global scale. However, despite their focus on conflict in for-
eign locations, most coding efforts have the limitation of exclusively relying on text
written in English. This paper advances the frontier of computerized event coding
by presenting the CAMEO Verb Translation Application (VTA), a new technology
to enable the translation of CAMEO verb dictionaries into non-English languages.
CAMEO VTA integrates automated translation with human-in-the-loop features to
increase the validity of the translation task. This development opens the possibility
of expanding the leverage of CAMEO in a systematic and efficient manner. The
first application of this tool focuses on translating CAMEO into Spanish.

Clionadh Raleigh

Dr. Raleigh is Professor of Political Geography and Conflict at the University of
Sussex. Raleigh previously served as a Lecturer in the Department of Political Sci-
ence at Trinity College, Dublin and an external researcher at the Peace Research
Institute Oslo (PRIO). Raleigh is a political geographer, and works mainly on con-
flict, governance, and the social consequences of climate change in sub-Saharan
Africa. Raleigh directs the Armed Conflict Location and Event Data (ACLED)
project, which tracks local conflict events in fifty countries.

Website

Key topics: armed conflict and political
violence; human-coded data; local
sources.

Selected Projects:

Armed Conflict Location and Event Data (ACLED)

The Armed Conflict Location & Event Data Project (ACLED) is a disaggregated
data collection, analysis, and crisis mapping project. ACLED collects the dates,
actors, locations, fatalities, and types of all reported political violence and protest
events across Africa, the Middle East, Latin America & the Caribbean, East Asia,
South Asia, Southeast Asia, Central Asia & the Caucasus, Europe, and the United
States of America. The ACLED team conducts analysis to describe, explore, and
test conflict scenarios, and makes both data and analysis open for free use by the
public.

Details and Data
Introductory Article

**Comparing Conflict Data: Simarlities and Differences across Conflict Datasets
(with Roudabeh Kishi)**

This report compares major human-coded (ACLED, UCDP-GED, GTD) and
machine-coded (ICEWS, GDELT, Phoenix) conflict datasets. The results show
how collection mandates, coding rules, and sourcing methods produce drasti-
cally pictures of conflict around the world, even when tracking ostensibly similar
phenomena.

Report

Idean Salehyan

Dr. Salehyan is Professor of Political Science at the University of North Texas and the co-Director of the Social Conflict in Africa Database project. Salehyan's research is often related to international & civil conflict, international migration, and politics the environment. In addition, Salehyan teaches courses on international relations, civil war, ethnic politics, & the politics of immigration.

Website

Key topics: social conflict; contentious actors; human-coded data.

Selected Projects:

Social Conflict in Africa Database (SCAD, with Cullen Hendrix)

Details and Data
Introductory Article

The Social Conflict Analysis Database (SCAD) includes protests, riots, strikes, inter-communal conflict, government violence against civilians, and other forms of social conflict not systematically tracked in other conflict datasets. SCAD currently includes information social conflicts from 1990-2017, covering all of Africa and now also Mexico, Central America, and the Caribbean.

Social Conflict Analysis Database-Organizational Properties (SCAD-OPs, with Ayal Feinberg and Kelsey Ann Naughton)

Details and Data
Introductory Article

The purpose of the Social Conflict Analysis Database—Organizational Properties (SCAD-OPs) is to provide more detail about the actors and targets listed in the original SCAD data; allow researchers to track the activities of a group over time; and ascertain relationships between groups. We focus on countries with a significant number of SCAD events (specifically, Mexico, Haiti, Dominican Republic, Egypt, Nigeria, Kenya, and South Africa), although coverage may expand to additional countries in the future. The data cover the time period from 1990-2016.

Organizers (Sooyeon Kang & Trey Billing)

Sooyeon Kang is a postdoctoral fellow at The Ohio State University's Mershon Center for International Security Studies. Kang studies mass mobilization, political violence, and political psychology. With Erica Chenoweth, Kang has developed the Major Episodes of Contention (MEC) dataset.

Website

Trey Billing is a postdoctoral fellow at The Ohio State University's Mershon Center for International Security Studies. Billing's research focuses on the dynamics and consequences of conflict, as well as the application of tools from machine learning and causal inference.

Website