VOLUME XVI, ISSUE 4 SPRING 2022 ISSUE

THE POLITICAL ECONOMIST

NEWSLETTER OF THE SECTION ON POLITICAL ECONOMY, AMERICAN POLITICAL SCIENCE ASSOCIATION

CO-EDITORS: CRISTINA BODEA, ANDREW KERNER, SHAHRYAR MINHAS, MICHIGAN STATE UNIVERSITY

From the Editors

Cristina Bodea, Andrew Kerner, Shahryar Minhas (Michigan State University)

We begin this edition of the Political Economist with a commemoration of **Frances McCall Rosenbluth**, written by **Thomas Pepinsky**. Frances had an enormous and enormously positive impact on the field. That is especially true for her students, including Tom. We are immensely grateful to Tom for sharing his tribute and allowing us to publish it here.

The remainder of this issue is dedicated to data. It features four articles focused new and non-traditional uses of data in political science research. These articles highlight the kinds of new data sources that political economists are utilizing, and the various challenges and opportunities that come along with them. Each essay speaks to a unique methodological approach, but they all suggest that we continue to think creatively about the data sources that we work with. **Jennifer Larson** writes on the challenges of preparing social network data from surveys, and discusses a process that scholars can use when doing so. **Scott Cook and David Fortunato** have a novel take on the use of more traditional data. They warn that data made available by governmental bodies exist downstream from a policy choice to collect and disseminate that data. Government-produced data are never truly apolitical, they argue, and analysts must consider the potential for manipulation and selection bias that are inherent to the data production process. A final pair of feature essays by **Francisco Cantú & Michelle Torres** and **Zachary Steinert-Threlkeld** discuss the exciting new frontier of images as data sources. They argue that for greater utilization of images as data, and provide a user-guide for doing so.

Finally, we want to direct readers to the new [Dataset Profiles] section on the [Political Economist website]. Dataset Profiles features links to new political economy datasets, along with short essays from the relevant authors describing the data, how they have used them, and how they might be used by others. We intend it to be a user-friendly promotional tool for new political economy data, and a vehicle to fuel innovative new research. If you are interested in discussing your work in Dataset Profiles please email Ha Eun Choi at choiha3@msu.edu or any of the other editors.

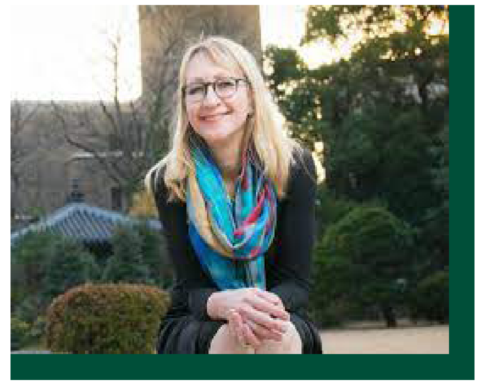
What's Inside This Issue:

In Commemoration

Frances McCall Rosenbluth

Thomas Pepinsky (Cornell University)

Frances McCall Rosenbluth, the Damon Wells Professor of political science at Yale University, died on November 20 after a battle with cancer. She was a prolific comparative political economist with enormous range, and a dedicated teacher. Frances was also a Chair of APSA's Political Economy section, and a leader in the discipline who touched the lives of countless researchers and students.



Many of us are saddened by her loss. Here, in appreciation of her life and career, I highlight some of her most important scholarly contributions to the field of political economy. My hope is that those who knew Frances will remember the brilliance in her work, and that those who did not are inspired to learn from her impressive record of top-notch political economy research.

Positive Political Economy and the Japanese Economic Miracle

Frances began her career as a scholar of Japanese politics and economics, writing a dissertation and later first book, Financial Politics in Contemporary Japan (Rosenbluth 1989), that addressed some of the core debates of the politics of economic policymaking. The question she addressed was, what explains financial reform in Japan? The prevailing academic orthodoxy of the late 1980s viewed the bureaucracy as the dominant actor within the Japanese political economy, a view that was central to the literature on the developmental state. In this account, financial reform happened because Japan's relatively autonomous bureaucracy decided that it was the best way to support Japan's economic transformation.

Frances adopted a different perspective: born in Osaka and fluent in Japanese, she could look very carefully at the internal politics of financial policymaking in ways that other Western scholars of the time could not. She argued that financial reform was driven by banks themselves. With close links to politicians from the ruling Liberal Democratic Party and to the powerful Ministry of Finance, powerful actors within the Japanese financial sector were able to adjust rules that they themselves desired to be bound by. This gave them an ability to compete in a changing global economy while also protecting their own firms' interests.

From the beginning of her career, Frances was thus a leading figure in introducing the tools of positive political economy into comparative politics, bringing political economy to bear on a key substantive policy problem. She invoked bureaucrats' career concerns rather than their ideological commitments; she brought the Stiglerian approach to regulation into conversation with theories of collective action; and she explained that the Ministry of Finance looked decisive because it nearly always made decisions that banks desired (and when MOF did not, the banks would use the tools at their disposal to overturn the MOF's decisions).

Frances's subsequent book, Japan's Political Marketplace (Ramseyer and Rosenbluth 1993), made this case more generally—it is a masterpiece of positive political economy that began from the premise that Japanese bureaucrats, politicians, and businesses followed the incentive structures created by the Japanese political system. In one memorable passage, she and her coauthor describe how their own fieldwork in Japan gave them a different view from other Japanologists, painting "a rather different picture—a picture of bureaucrats far less effective than the 'plan-rational' engineers who inhabit most political science monographs" (p. 141). Their approach took seriously how information and monitoring costs influenced the logic of delegation in a dominant-party system such as Japan. This is almost certainly the first book on East Asia's political economy to have explicitly argued that if we look only at policy outputs, a strong and independent bureaucracy is observationally equivalent with a politically subservient one.

The Political Economy of Electoral Institutions

Early in her career, Frances turned her formidable intellect to questions of party competition and institutional change. Her knowledge of the Japanese case helped, but this work established her as a comparativist whose theoretical understanding of how institutional rules affect politics would prove widely applicable for scholars of party politics and electoral rules.

One of the most distinctive features of Japanese politics is factionalism within the LDP, an otherwise hegemonic political party that has held power with only brief interruptions since 1955. Through the 1970s, factional bosses were responsible for the majority of fundraising for their factions, but as this began to change in the 1980s, it had implications for the relative power and prominence of the factions in the LDP. Frances showed that this change led to a form of convergence across factions, leading them to be less distinctive from one another electorally by the 1990s (Cox and Rosenbluth 1993). Her subsequent work on the political economy of factional politics in Japan looked at elite mobilization and turnout (Cox, Rosenbluth, and Thies 1998) and how electoral rules incentivize LDP members to join factions in the first place (Cox, Rosenbluth, and Thies 2000). The latter of these contributions is the first to have exploited variation across time in Japan's electoral rules, before and after the 1994 reform of Japan's electoral system.

These works are now canonical contributions to the political economy of electoral institutions. And at the same time, Frances was building a comparative research agenda on how party competition affects the public sector. The showpiece for this agenda is seminal article in which Frances showed that advanced democracies with more political parties (and hence coalition governments) have higher levels of government spending than those with fewer parties (Bawn and Rosenbluth 2006). Once again, this work demonstrates the power of positive political economy in comparative politics.

The Political Economy of Gender

Although her contemporaries remember her fundamental contributions to the political economy of electoral institutions and the political economy of Japanese politics, Frances is perhaps best known among early-career political economists today for her more recent pathbreaking research on the political economy of gender.

Frances argued that to understand gender inequality in capitalist economies, we must begin with the incentives that women face in a world in which the primary burden of childcare still falls primarily to them. This line of reasoning has a long pedigree in political economy, and one need look no further than Gary Becker's theoretical approach to the economics of the family, used to understand the intrahousehold division of labor, the decision to bear children, investing in children's education, and other topics. But Becker's approach to the family did not have much to say about politics, even though scholars of welfare regimes have long recognized that much of contemporary social policy comes down to how states recognize (or not) the work done by women within the household.

Frances provided us with a series of fundamental contributions to the political economy of gender that recognized that state institutions don't just recognize labor in the household, they also shape the opportunities that women have to select out of traditional family arrangements. Understanding how the possibility of divorce affects women's bargaining power within the family, as well as the potential variation between the interests of women and men, and Frances showed that the intrahousehold division of labor is most unequal in countries where specific-skills are more valuable than general skills and where the public sector does not respond by subsidizing childcare and women's employment (Iversen and Rosenbluth 2006). Rather provocatively, one conclusion from this article is that the intrahousehold division of labor would be least unequal either in liberal market economies like the United States or in specific-skill regimes with large public sectors like Norway. Japan—where Frances started her career as a political economist—stands out as uniquely unequal precisely because its economic system emphasizes firm-specific skills without compensating women for the time they must spend away from home in order to acquire those skills. Her landmark book Women, Work, and Politics (Iversen and Rosenbluth 2010) expanded this argument further, and was the basis of one of Yale's most popular undergraduate courses.

None of this work emphasized the inherent qualities of women as caregivers, or patriarchal norms that repress women's autonomy. Instead, Frances viewed women as economic actors within political systems that varied in the incentives and constraints that they placed on women and men. Her approach was relational and institutionalist. To understand, for example, why some countries have higher rates of women's in office than others (Iversen and Rosenbluth 2008), she considered how party systems incentivize candidate- versus party-centric relationships between politicians and their constituents. Candidate-centric systems require candidates to invest heavily in those relationships, and the time required to do so places women at a disadvantage relative to men (once again, given the unequal household division of labor), with the implication that candidate-centric systems will feature fewer women in politics. Not because women are bad at politics, or because voters and parties are inherently sexist (although some are), but because the institutional structure of party politics places women at a structural disadvantage. In one of her most recent and most prominent contributions, Frances showed that this logic also extends to the expectations that American voters have for candidates for office. American voters want to vote for women, but they also want to vote for candidates with children and a traditional marriage—exactly the social roles that place unequal burdens on women (Teele, Kalla, and Rosenbluth 2018).

A Mentor for Many

This review of Frances's major contributions to political economy doesn't come close to covering all of them. She continued to work on the comparative politics of banking regulation (Rosenbluth and Schaap 2003), on the role of warfare in democratic societies (Ferejohn and Rosenbluth 2016), and even looked far back in the archaeological record to identify the political determinants of wealth inequality using osteological data—that is, human bones (Boix and Rosenbluth 2014). And she contributed a wealth of other work on Japan, on democracy, on judicial politics, and on related topics.

In other memorial statements, I and others have written about how transformational Frances was as a mentor: all of the dissertations she supervised, all of the students she supported, and the absolutely incredible role model that she was for women (and men) in the profession. Here, I will offer a brief and more personal reflection on how Frances supported one political economist.

I don't recall ever talking about Japanese politics with Frances, or about the political economy of finance. I remember, instead, debating classic works of positive political economy during my first semester of grad school, her generosity in leading an independent study with me and two classmates who wanted to know even more about political economy, and her confidence that my research on Asia did not need to be peripheral to comparative politics. I think a lot of other

Rosenbluth (cont.)

Pepinsky

students had an experience like I had. But I didn't realize until very recently that I ended up following the path that she set out for us. As she did in so many aspects of her life, she led by example, and I am grateful that she was my teacher, my mentor, and my friend.

References

Bawn, Kathleen, and Frances Rosenbluth. 2006. "Short versus Long Coalitions: Electoral Accountability and the Size of the Public Sector." American Journal of Political Science 50 (2):251-265. doi: https://doi.org/10.1111/j.1540-5907.2006.00182.x.

Boix, Carles, and Frances Rosenbluth. 2014. "Bones of Contention: The Political Economy of Height Inequality." American Political Science Review 108 (1):1-22. doi: 10.1017/S0003055413000555.

Cox, Gary W., and Frances Rosenbluth. 1993. "The Electoral Fortunes of Legislative Factions in Japan." American Political Science Review 87 (3):577-589. doi: 10.2307/2938737.

Cox, Gary W., Frances M. Rosenbluth, and Michael F. Thies. 1998. "Mobilization, Social Networks, and Turnout: Evidence from Japan." World Politics 50 (3):447-474. doi: 10.1017/S0043887100012879.

Cox, Gary W., Frances M. Rosenbluth, and Michael F. Thies. 2000. "Electoral Rules, Career Ambitions, and Party Structure: Comparing Factions in Japan's Upper and Lower Houses." American Journal of Political Science 44 (1):115-122. doi: 10.2307/2669297.

Ferejohn, John, and Frances McCall Rosenbluth. 2016. Forged Through Fire: War, Peace, and the Democratic Bargain. New York: W.W. Norton.
Iversen, Torben, and Frances Rosenbluth. 2006. "The Political Economy of Gender: Explaining Cross-National Variation in the Gender Division of Labor and the Gender Voting Gap." American Journal of Political Science 50 (1):1-19. doi: https://doi.org/10.1111/j.1540-5907.2006.00166.x.

Iversen, Torben, and Frances Rosenbluth. 2008. "Work and Power: The Connection Between Female Labor Force Participation and Female Political Representation." Annual Review of Political Science 11 (1):479-495. doi: 10.1146/annurev.polisci.11.053106.151342.

Iversen, Torben, and Frances McCall Rosenbluth. 2010. Women, Work, and Politics: The Political Economy of Gender Inequality. New Haven: Yale University Press.

Ramseyer, J. Mark, and Frances M. Rosenbluth. 1993. Japan's Political Marketplace. Cambridge: Harvard University Press.

Rosenbluth, Frances M. 1989. Financial Politics in Contemporary Japan. Ithaca: Cornell University Press.

Rosenbluth, Frances, and Ross Schaap. 2003. "The Domestic Politics of Banking Regulation." International Organization 57 (2):307-336. doi: 10.1017/S0020818303572034.

Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." American Political Science Review 112 (3):525-541. doi: 10.1017/S0003055418000217.

Preparing Social Network Data

Jennifer Larson (Vanderbilt University)

Data that contain rich information about the relationships between people are becoming increasingly prevalent in political science research (Atwell and Nathan (2021), Arias et al. (2019), Ferrali et al. (2020), Larson and Lewis (2017), Larson, Lewis and Rodríguez (Forthcoming)). Social network data let scholars investigate topics such as how people spur one another to protest, how friends reinforce political attitudes, how insurgent groups recruit, and how people come to learn new information that shapes their political behavior, to name just a few.

There are a variety of ways to collect the relationships, or "links" needed in social network data. Interactions on media leave traces that can be collected into a dataset recording who follows whom or who retweets whom. Public biographies reveal shared experiences like attending the same university or growing up in the same city that can be used to connect observations. Actions in the public eye such as serving on the same committees or co-sponsoring legislation can be recorded as the links between actors. And when the links of interest are less observable, perhaps because the interactions are private or the relationships are less tangible, scholars typically turn to surveys as a means of acquiring information about the links. Surveys can reveal harder-to-observe friendship, trusted confidants, and the people someone regularly encounters in the tasks of daily life.

In a name-generator survey technique, respondents are asked to name a certain number of people who meet some criterion. "Think about all the people who do not live in your household with whom you shared a meal last month. Please name up to five of them." Each additional network type would have its own name-generator question; surveys often include more than one. These survey modules lead to a dataset with a row for each respondent, as is customary, columns for every individual-level variable included, and columns for each name given in answer to each name-generator question. (For more on this style of data collection, see Larson and Lewis (2020).) Network data collected in this way results in each respondent potentially offering many names of other people.

The trick is that the researcher needs to turn all of these names– the respondents and the named names– into a network (see Fig 1). This requires making two matches: if respondent A lists respondent B in response to a network question, the researcher needs to identify that the listed name is in fact respondent B. Likewise, if respondent C and respondent D both mean to list the same person in response to a network question, the names they list need to in fact indicate the same person. In a perfect data collection world, all names in the data would appear completely, in a single format, and spelled a single way. In practice, names get entered with substantial variation. Consider a few possible sources:

- Respondents spell others' names incorrectly.
- Respondents do not know the spelling of others' names.
- Enumerators enter names with some error.
- Respondents tell you their own name in one format, say first, middle, and last name, but then refer to others using a different format, perhaps just their first and last name.

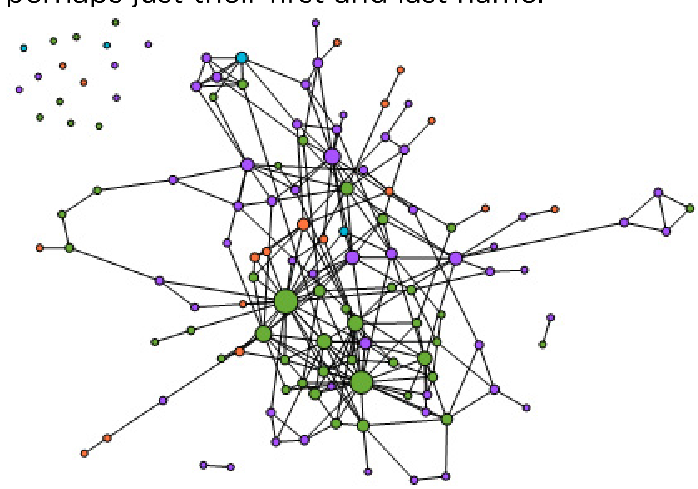


Figure 1: A social network measured with a name-generator survey.

This issue of matching is potentially a large one. Consider the respondents in the upper left of Figure 1, who appear as floating nodes with no links. In the survey that led to this network, these respondents did offer names in response to the networks questions; those names just did not match any of the names of the other respondents. That could be because they did mean different people, in which case this view of the network is correct. Or it could be that the enumerator spelled the name they said differently, or introduced a typo, or included more or fewer names in the full name. The problem is that even a few missed links could substantially change our view of a network.

In my experience, the names collected from surveys can be surprisingly messy. Fortunately, there are steps that can help reduce the mess and increase confidence in the measured networks. Additionally, anticipating the need to clean names in network data can pay off substantially. There are simple questions that can be added to a survey that help calibrate choices that must be made when cleaning the names.

Cleaning names in network data

The method laid out here for cleaning network data proceeds in the following steps:

1. Manually inspect lists of names in the data to look for particular anomalies or issues that should be addressed.
2. Pre-process all names in the data.
3. Identify a reference set of names, typically the egos or their households.
4. Pick a threshold of similarity above which two names count as the same.
5. Loop over all names in all name variables in the dataset (other than the variables establishing the reference set), compare them to the reference set names, and replace them with the reference set name when their similarity exceeds the threshold.
6. Use calibration variables to assess the likely error introduced by that threshold. Anticipating this step is key to designing robust data collection.
7. Repeat steps 5 and 6 with other thresholds.
8. Select the best threshold, create master dataset using it to clean the names.
9. Select another one or two thresholds, create datasets to use in robustness checks to determine if analyses done using the data are sensitive to the cleaning.

Manual inspection, pre-processing

Looking at the names in the dataset all together is the best way to spot issues and make informed decisions about the automated cleaning. Did people usually list first and last names? Did they include anything extra such as titles, nicknames, or parentheticals? Surveys that ask respondents to name "full names" may still return a variety of formats and names with extra characters. Ultimately, creating the network will assume that names that are meant to refer to the same person appear in exactly the same form in the data. This first step will help identify things that may prevent perfect matches.

Once you have a handle on the kinds of extra characters that may appear, you can treat the names as text data and pre-process them to share a common form. One form that is particularly useful for the automated steps that follow is: all capital letters, without punctuation free, with a single space separating the multiple names of a full name, and with the multiple names stored in alphabetical order within a full name.

Selecting a reference set

Ideally your data contain a set of names that you are particularly interested in. These might be the respondents that you spoke with. Each of these respondents might have listed many other names in response to the survey's network questions, so many other names may appear in the data. But the respondents form a set of names that you know are different from one another (presumably you did not conduct the survey with the same person twice), and whether respondents listed any of *these* names in their response to networks questions is likely of particular interest. In some cases, researchers want to construct household networks instead of individual networks, and so each respondent is asked to list the other members of her household. Ultimately the network will be constructed with households as the nodes, and links connect two households according to some rule (such as that a link is present between two households if anyone in one household lists any member of the other household). In this case, selecting the set of respondents as well as the people they list as their household members would be a sensible reference set.

Match names based on a threshold of similarity

The next step is to determine whether any of the names listed in response to the network questions should be considered the same as a name in the reference set. In perfectly collected and recorded network data, there would be no ambiguity. If there is a "Jerry Lopez" in the reference set, then everyone who meant to list Jerry Lopez in response to a network question would write (or say and the enumerator would write) "Jerry Lopez". A perfect match. But that intention could also show up in the data as Jery Lopez, Jerri Lopez, Jerry Q. Lopez, and so on. String difference metrics are a useful way to determine how similar two names are. The Levenshtein distance metric calculates the minimum number of single-character insertions, deletions, or substitutions that would change one version of the name into another. "Jerry Lopez" is Levenshtein distance 1 from "Jery Lopez", 1 from "Jerri Lopez", and 3 from "Jerry Q. Lopez" (2 after pre-processing removes the period).

We can clean all names in a dataset by picking a threshold (say Levenshtein distance 1), looping over all non-reference set name variables' names, calculating each name's Levenshtein distance to all names in the reference set, and replacing it with the name in the reference set from which it is within the threshold distance. (The r function adist() is useful here). This method preserves the names in the reference set, and adjusts other names listed to exactly match names in the reference set if they are similar enough.

Calibrating the threshold

Cleaning names in this way risks adding matches to the reference set where none were intended. The choice of the threshold does not need to be made blindly. The initial inspection of the names, especially if viewed alphabetically, will give some sense of whether these few-character-differences are a likely problem, as well as the rough length of names (relevant because any two names would count as the same if the Levenshtein distance threshold is high enough; the shorter the names, the lower that threshold could be).

Moreover, simple questions built into the survey can help assess the quality of the cleaning. For example, asking a question about each name offered such as "do they live in your village?" is quick, and gives a lot of information to aid calibration. After cleaning with a particular threshold, count the number of unique names in the data set for which respondents said that person lives in the village. If the number exceeds the population of the village, the threshold is likely too low. (Names that should probably refer to the same person are still appearing differently).

Additionally, this measure can give a sense of over-cleaning as well. How many names are there now that the respondent said do not live in this village, but that this round of cleaning matched to one of the respondents anyway? Too many means the threshold is probably too high– the cleaning is imposing matches where it should not. Noting the number of names that are replaced by the cleaning can help inform judgment about over-cleaning as well, and manually inspecting the changed name is always a good idea.

Retaining the results of other thresholds for robustness checks

Being certain that any one network is exactly correct, be it the uncleaned original or one cleaned to a certain similarity threshold, is in most cases impossible. Keeping datasets cleaned at different thresholds can be an important safeguard once performing the analyses for which the network was ultimately collected. Multiple datasets cleaned in different ways allows for sensitivity analysis. Replicating key findings on datasets cleaned differently can help improve confidence. For instance, if a look at the names suggested that cleaning was needed, the threshold chosen resulted in a majority of named names being changed to match the reference set, but analyses tell the same story when using data cleaned with the next highest and next lowest threshold, then we would have greater confidence in the results. On the other hand, knife-edge results that only appear in one aggressively cleaned dataset but not others should give us greater pause.

Getting the Network Right

Of course, this only addresses one source of potential measurement error in network data. Issues of similar names that are in fact different people, repeated names, and nicknames that are written nothing like the full name will not be addressed by this round of cleaning. However, the same principles of thinking in advance about discriminating questions that could be added to the survey can help. If multiple names are likely going to be an issue, could the survey also collect a respondent's guess of the person's age? Their mother's name? The number of children they have? A little qualitative probing in advance for good, discriminating questions and remembering to include them in the survey can go a long way to protect against mismeasured networks.

References

Arias, Eric, Pablo Balán, Horacio Larreguy, John Marshall, and Pablo Querubín. "Information provision, voter coordination, and electoral accountability: Evidence from Mexican social networks." American Political Science Review 113, no. 2 (2019): 475-498.

Atwell, Paul, and Noah L. Nathan. "Channels for influence or maps of behavior? a field experiment on social networks and cooperation." American Journal of Political Science (2021): Firstview.

Ferrali, Romain, Guy Grossman, Melina R. Platas, and Jonathan Rodden. "It takes a village: peer effects and externalities in technology adoption." American Journal of Political Science 64.3 (2020): 536-553.

Larson, Jennifer M., and Janet I. Lewis. "Ethnic networks." American Journal of Political Science 61.2 (2017): 350-364.

Larson, Jennifer M., and Janet I. Lewis. "Measuring networks in the field." Political Science Research and Methods 8.1 (2020): 123-135.

Larson, Jennifer M., Janet I. Lewis, and Pedro Rodríguez. "From Chatter to Action: How Social Networks Inform and Motivate in Rural Uganda." British Journal of Political Science. Forthcoming.

What Convolutional Neural Networks can (and cannot) do for social scientists

Francisco Cantu (University of Houston) & Michelle Torres (Rice University)

In the past years, technological advancements have allowed us to access and analyze large amounts of unstructured data such as texts, speech, and images. These tools have allowed social scientists to automate the classification and extraction of information from large pools of pictures and videos. As a result, researchers use this information to answer relevant questions regarding social movements[1],decision making[2], political communication[3], campaigns[4], or attitude formation[5].

One of the most popular models for analyzing visual content is the Convolutional Neural Network (CNN or ConvNet)[6]. This type of neural network aims to emulate how humans learn to process visual information[7]. Infants learn to recognize an object by subconsciously assimilating its distinctive properties. For example, after being exposed multiple times to images of a cat (either by observing a real one or photos of it), they learn that a "cat" has paws, two ears, and whiskers. The infant might not know those features in isolation or their particular names, but the neurons in her brain process the distinct parts of such elements (e.g. color, texture, shape), allowing the recognition and identification of cats in multiple circumstances.

Similarly, a CNN is composed of layers with "filters" looking for different features of an image. These filters are organized into layers with different levels of complexity. Initial layers look for the existence of horizontal and vertical lines. Intermediate layers may look for basic shapes like circles or objects. Finally, "deeper" layers are in charge of processing the information output by the previous layers to then scan for more complex features like color combinations, textures, and objects[8]. Using multiple labeled examples assigned by the researcher, the model learns the relevant features associated with a category of interest. Analogous to a child interacting with her cat over weeks, the computer is exposed to millions of frames of a cat but in a matter of minutes. The ultimate goal for the CNN during this training process is to be able to identify all the cats in an out-of-sample picture or to distinguish between a cat or a dog.

The structure and objectives achieved by CNNs have allowed scientists from multiple fields to automate the analysis of millions of images faster and more efficiently than ever. However, social scientists should be aware of the strengths and limitations of CNNs to achieve their substantive goals. In this piece, we discuss a few of the characteristics of CNNs that affect their impact and potential. We provide examples of concrete tasks and objectives that this model executes with efficiency and accuracy, and we also flag certain areas where

---

1 Andreu Casas and Nora Webb Williams, "Images that matter: Online protests and the mobilizing role of pictures," Political Research Quarterly 72, no. 2 (2019): 360-375; Han Zhang and Jennifer Pan, "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media," Sociological Methodology 49, no. 1 (2019): 1-57;
Donghyeon Won, Zachary C Steinert-Threlkeld, and Jungseock Joo, "Protest activity detection and perceived violence estimation from social media images," in Proceedings of the 25th ACM international conference on Multimedia (ACM, 2017), 786-794.

2 Bryce J Dietrich, Matthew Hayes, and Diana O'Brien, "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech," American Political Science Review 113, no. 4 (2019): 941-962; Justin Grimmer, Solomon Messing, and Sean J Westwood, "How words and money cultivate a personal vote: The elect of legislator credit claiming on constituent credit allocation," American Political Science Review 106, no. 4 (2012): 703-719.

3 Justin Grimmer, Representational style in Congress: What legislators say and why it matters (Cambridge University Press, 2013).

4 Markus Neumann, "Fair and Balanced? News Media Bias in the Photographic Coverage of the 2016 U.S. Presidential Election" (Working paper, 2019);
Sebastian Stier et al., "Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter," Political communication 35, no. 1 (2018): 50-74;
Constantine Boussalis et al., "Gender, candidate emotional expression, and voter reactions during televised debates," American Political Science Review 115, no. 4 (2021): 1242-1257.

5 Michelle Torres, "Framing a Protest: Determinants and Eects of Visual Frames" (Working Paper, 2018).

6 Yann LeCun and Yoshua Bengio, "Convolutional networks for images, speech, and time series," in The Handbook of Brain Theory and neural networks, 2nd ed., ed. Michael Arbib (Cambridge, MA, 2003), 276-280;
Yann LeCun et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE 86, no. 11 (1998): 2278-2324.

7 For a richer discussion regarding the parallels and synergies between brain functioning and computer vision models, see David Marr, Vision (Cambridge, MA: MIT Press, 2010).

8 Michelle Torres and Francisco Cantu, "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data," Political Analysis 30, no. 1 (2022): 113-131.

CNNs (cont.)

Cantu & Torres

extra caution is necessary. As with many other computational and statistical tools, great power comes with great responsibility. A proper understanding of the logic, implementation, strengths, and flaws of CNNs allows a more impactful use of them.

What CNNs can (and cannot) do

Classification and identification of factual concepts

A very simplified intuition behind CNNs consists of exposing the computer to multiple images of a label (or labels) of interest so the model learns what combination of pixels and patterns among them are more strongly associated with the different categories. The canonical example is the recognition of handwritten numbers[9]. After feeding a CNN with several samples of pictures of handwritten digits, the model learns that most "1"s *generally* have a vertical straight line, that a "0" is a round closed oval, and that an "8" has two stacked closed areas. A CNN trained on traditional datasets[10] reaches very high levels of accuracy and prediction, even with relatively simple architectures.

A crucial aspect behind the effectiveness of CNNs predicting handwritten numbers is the *factual* nature of the data: for most of the digits under analysis, there is a clear ground truth that human coders can classify and corroborate. In other words, the labels attached to the images of interest are not subjective or dependent on the perceptions and biases of the viewer. The logic behind this is that computers and models learn a concept better when there are consistent and unambiguous samples of it in the training set.

When the data is consistent and unambiguous, CNNs are perfect models for labeling and classifying its content. A few relevant examples that fulfill these conditions include the identification of public figures, the tracking of police officers in a public event, the detection of objects, like flags, in social media pictures, and the classification of the activities depicted in pictures of interactions between international leaders such as shaking hands, document signing, or a conversation. When it comes to completing these goals, the CNN acts as an efficient research assistant who is immune to exhaustion, boredom, and distractions. It allows to process a large number of images in a very short period of time and make minimal to zero mistakes at all.

However, the ability of CNNs to classify and "recognize" concepts is sometimes overestimated and also misunderstood. Just as in many cases involving computational tools, if humans cannot reliably classify a specific label, the computer will not be able to do so either. This is, if there is variability in what human coders perceive due to cognitive biases, characteristics of the observer, confusion about the label, or the abstraction of the concept, then a CNN will struggle to find clear patterns and thus to perform the classification task accurately.

An example of this case is the coding of emotions that an image "evokes"[11]. Social scientists are interested in understanding not only what emotions political images trigger, but also in *how* the specific characteristics of a photo and its associated metadata (e.g. creator, publisher, context) influence emotional reactions that in turn catalyze behavior or attitude formation processes. The question is of utter importance to understanding the impact of images. However, training a model to recognize these evoked emotions is a non-trivial task.

Building a training set demands "coherent definitions of concepts for particular applications," and it implies two crucial steps: the creation of a coding scheme, and an appropriate sampling of documents that human coders will process and classify[12]. The creation of a coding scheme is an iterative process where coders detect ambiguities and confusing examples, the researcher spots disagreement between coders and designs rules to avoid different interpretations, and all the actors involved reach an agreement on the definition of each of the labels of interest.

For example, a researcher is interested in finding images that trigger disgust to study whether this emotion influences respondents' opinions about the pandemic even in those cases where the emotion is not related to the outcome of interest[13]. To achieve this, she provides several students with images of different animals

9 Yann LeCun et al., "Backpropagation applied to handwritten zip code recognition," Neural computation 1, no. 4 (1989): 541-551.

10 The Modied National Institute of Standards and Technology database (MNIST) database is the most popular dataset to train and test models for handwritten digits recognition. It is composed of 60,000 digits for training, and 10,000 for testing Yann LeCun, "The MNIST database of handwritten digits," http://yann. lecun. com/exdb/mnist/,1998,

11 Casas and Webb Williams, "Images that matter: Online protests and the mobilizing role of pictures."

12 Justin Grimmer and Brandon Stewart, "Text ad Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," Political Analysis 21, no. 3 (2013): 267-297.

13 For a similar study, see Scott Cliord and Jennifer Jerit, "Disgust, anxiety, and political learning in the face of threat," American Journal of Political Science 62, no. 2 (2018): 266-279.

CNNs (cont.)

Cantu & Torres

and asks them to code whether each of them makes them feel disgusted. Although 90% of the students reacted negatively to rats and positively to dogs, the sample was split when it came to butterflies: half of the students think they are beautiful and felt happy after seeing them, while the other half find them creepy and felt disgusted after being exposed to them. This is a cartoon example of how a depiction of a simple object can trigger different emotional reactions. It gets more consequential and complex when dealing with political visual content, where the ideologies and personal characteristics of the audience filter how they interpret it[14]. A picture of a young man being arrested during a Black Lives Matter protest can evoke opposite emotions in viewers. A conservative one who favors law and order might show *relief* that the system is punishing agitators, while a progressive activist seeking social justice might feel *anger* after seeing this action as unfair. Because of the subjectivity and ambiguity of the trait of interest, it is hard to define a ground truth and therefore a consistent concept that the CNN can learn to then complete a large-scale classification of images according to this trait. Thus, researchers should be aware of their concepts of interest and evaluate whether a CNN will be appropriate for their analysis.

Data collection and processing

The goal of classifying or identifying particular concepts of interest in images is a common way of "collecting" a variable of interest. However, the abilities of CNNs to accurately recognize factual data turn them into powerful tools for other less traditional types of data collection as well. Social scientists study questions that rely on the information contained in maps, historic registries, and other types of archival sources. For example, the ability to combine individual information on millions of subjects from historical censuses and Confederate rosters has allowed researchers to better understand how slavery shaped the choice to fight for the Confederacy[15]. Similar exercises in which researchers digitized historical electoral results, collected and classified electoral tallies, and reviewed early 20th-century personal notes and hand-written letters from African American women, have illuminated questions about the effect of gender-separated voting[16], the existence of electoral fraud[17], and the factors shaping the lives and legacies of women[18].

While the ability to analyze new and large sources of data has dramatically expanded in recent years, the fact remains that much of the raw data necessary for better understanding individual and collective political behavior is stored as physical hardcopies with handwritten information and other elements that researchers need to extract, recognize, and process to make it machine-readable. In most cases, these tasks require human coders to manually extract and label the data. Although this approach might be appropriate when dealing with a small sample of documents, it becomes unsustainable when there are millions of pages to process.

CNNs can become excellent assistants for detecting handwritten digits and characters in an image of a document[19]; identifying complex layouts, such as those in newspapers, for further extraction of images, ads, and text[20]; recognizing relevant elements like seals or stamps, and others. Beyond the suitability of the data, an important advantage of the application of CNNs to information retrieval tasks is that there already exist several models and architectures designed to perform them. This heavily reduces the need for intensive and costly training exercises involving a large number of images.

In most cases, researchers are still required to customize existing architectures to their particular data and needs, but this step is considerably less demanding than training a CNN from scratch. In fact, most researchers interested in visual analysis use *transfer learning* to achieve their specific objectives in an efficient way. This approach uses the features that a previously trained CNN has already learned from an extensive pool of images. These features are then "tuned" to a more specific set of labels. In contrast to full training, transfer

14 Nora Webb Williamns et al., "When Republicans See Red but Democrats Feel Blue: Why Labeler Characteristic Bias Matters for Image Analysis" (2021).

15 Andrew B Hall, Connor Hu, and Shiro Kuriwaki, "Wealth, slave ownership, and fighting for the confederacy: An empirical study of the American civil war," American Political Science Review 113, no. 3 (2019): 658-673.

16 Jonathan Homola, "The Political Consequences of Group-Based Identities" (2018).

17 Francisco Cantu, "The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election," American Political Science Review 113, no. 3 (2019): 710-726.

18 Ula Taylor, "Women in the documents: Thoughts on uncovering the personal, political, and professional," Journal of Women's History 20, no. 1 (2008): 187-196.

19 Torres and Cantu, "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data."

20 Zejiang Shen et al., "LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis," arXiv preprint arXiv:2103.15348, 2021.

CNNs (cont.)

Cantu & Torres

learning requires a much smaller set of images providing examples of the outcome of interest[21]. An example of this process is to take a CNN trained on images of scenes, landscapes, and daily activities, "freeze" the last layers that output the labels of those concepts, and instead use a new sample of images of protests to teach the model about the specific characteristics of these social movements so it can then recognize these events. The first layers already determined and learned the features that distinguish forests from cities, and people swimming from having coffee. Thus, the new training set will help to distinguish some of the images according to the characteristics of a protest.
The "transferring" of knowledge from the first layers of a CNN, built based on data from one domain (pictures of scenes), to the final layers of a network tends to yield optimal results about the specific, but at best *vaguely* related, domain of interest (pictures of protest)[22]. In fact, there are cases where that knowledge can be detrimental for the desired classification goal[23], if the source and target domains are not sufficiently related[24].
Thus, the use of transfer learning has another advantage for tasks related to information retrieval and recognition: a potential increase in the accuracy of the classifications. Since many of the models tailored to these objectives are trained on large amounts of data that resemble or share many of the features of the target data, we can expect the learning from the source data to be more impactful and efficient in the final classification. For a practical example and extended explanation of this, see our published article on Convolutional Neural Networks[25].

What is next?

In this piece, we are interested in providing some examples of ways in which social scientists can maximize the advantages of CNNs. These computational tools are powerful models that can efficiently label and classify millions of images (and pieces of images) in record times. However, by understanding the way in which CNNs operate and yield results, researchers should be cautious about the applications for which they use them.
The accessibility of both data and analysis tools opens new exciting avenues of research and allows for the analysis of new and understudied questions. However, they also demand from us a deep understanding of their strengths and flaws, as well as their scope and impact. In the case of CNNs, their increasing use also calls for more careful validation exercises and transparency in the report of limits and mistakes that we encounter. It also motivates a more comprehensive analysis of the ways in which we can make them more impactful for social scientists, for example by bringing to the table concepts like interpretation and inference. This involves creating and sharing datasets that are strictly related to social science domains, as well as developing models and analysis tools that consider the particular objectives and data that social scientists have.

21 For a more detailed explanation, as well as notes on the implementation of this concept in the social sciences, please see Nora Webb Williams, Andreu Casas, and John D. Wilkerson, Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classication (Cambridge University Press, Forthcoming).

22 Maxime Oquab et al., "Learning and transferring mid-level image representations using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition (2014), 1717-1724.

23 Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang, "A survey of transfer learning," Journal of Big Data 3, no. 1 (2016): 1-40.

24 Michael T Rosenstein et al., "To transfer or not to transfer," in NIPS 2005 workshop on transfer learning, vol. 898 (2005), 1-4.

25 Torres and Cantu, "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data."

References

Boussalis, Constantine, Travis G Coan, Mirya R Holman, and Stefan Muller. "Gender, candidate emotional expression, and voter reactions during televised debates." American Political Science Review 115, no. 4 (2021): 1242–1257.

Cantu´, Francisco. "The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election." American Political Science Review 113, no. 3 (2019): 710–726.

Casas, Andreu, and Nora Webb Williams. "Images that matter: Online protests and the mobilizing role of pictures." Political Research Quarterly 72, no. 2 (2019): 360–375.

Clifford, Scott, and Jennifer Jerit. "Disgust, anxiety, and political learning in the face of threat." American Journal of Political Science 62, no. 2 (2018): 266–279.

Dietrich, Bryce J, Matthew Hayes, and Diana O'Brien. "Pitch Perfect: Vocal Pitch and the Emo- tional Intensity of Congressional Speech." American Political Science Review 113, no. 4 (2019): 941–962.

Grimmer, Justin. Representational style in Congress: What legislators say and why it matters.
Cambridge University Press, 2013.

CNNs (cont.)

Cantu & Torres

Grimmer, Justin, Solomon Messing, and Sean J Westwood. “How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation.” American Political Science Review 106, no. 4 (2012): 703–719.

Grimmer, Justin, and Brandon Stewart. “Text ad Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” Political Analysis 21, no. 3 (2013): 267–297.

Hall, Andrew B, Connor Huff, and Shiro Kuriwaki. “Wealth, slaveownership, and fighting for the confederacy: An empirical study of the American civil war.” American Political Science Review 113, no. 3 (2019): 658–673.

Homola, Jonathan. “The Political Consequences of Group-Based Identities.” 2018.
LeCun, Yann. “The MNIST database of handwritten digits.” http://yann. lecun. com/ exdb/mnist/, 1998.

LeCun, Yann, and Yoshua Bengio. “Convolutional networks for images, speech, and time series.” In The handbook of brain theory and neural networks, 2nd ed., edited by Michael Arbib, 276–280. Cambridge, MA, 2003.

LeCun, Yann, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. “Backpropagation applied to handwritten zip code recognition.” Neural computation 1, no. 4 (1989): 541–551.

LeCun, Yann, L´eon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition.” Proceedings of the IEEE 86, no. 11 (1998): 2278–2324.
Marr, David. Vision. Cambridge, MA: MIT Press, 2010.

Neumann, Markus. “Fair and Balanced? News Media Bias in the Photographic Coverage of the 2016 U.S. Presidential Election.” Working paper, 2019.

Oquab, Maxime, Leon Bottou, Ivan Laptev, and Josef Sivic. “Learning and transferring mid-level image representations using convolutional neural networks.” In Proceedings of the IEEE conference on computer vision and pattern recognition, 1717–1724. 2014.

Rosenstein, Michael T, Zvika Marx, Leslie Pack Kaelbling, and Thomas G Dietterich. “To transfer or not to transfer.” In NIPS 2005 workshop on transfer learning, 898:1–4. 2005.

Shen, Zejiang, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, and Weining Li. “LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis.” arXiv preprint arXiv:2103.15348, 2021.

Stier, Sebastian, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. “Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter.” Political communication 35, no. 1 (2018): 50–74.

Taylor, Ula. “Women in the documents: Thoughts on uncovering the personal, political, and pro- fessional.” Journal of Women’s History 20, no. 1 (2008): 187–196.

Torres, Michelle. “Framing a Protest: Determinants and Effects of Visual Frames.” Working Paper, 2018.

Torres, Michelle, and Francisco Cantu´. “Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data.” Political Analysis 30, no. 1 (2022): 113–131.

Webb Williamns, Nora, Andreu Casas, Kevin Aslett, and John Wilkerson. “When Republicans See Red but Democrats Feel Blue: Why Labeler Characteristic Bias Matters for Image Analysis.” 2021.

Webb Williams, Nora, Andreu Casas, and John D. Wilkerson. Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification. Cambridge University Press, Forthcoming.

Weiss, Karl, Taghi M Khoshgoftaar, and DingDing Wang. “A survey of transfer learning.” Journal of Big data 3, no. 1 (2016): 1–40.

Won, Donghyeon, Zachary C Steinert-Threlkeld, and Jungseock Joo. “Protest activity detection and perceived violence estimation from social media images.” In Proceedings of the 25th ACM international conference on Multimedia, 786–794. ACM, 2017.

Zhang, Han, and Jennifer Pan. “CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media.” Sociological Methodology 49, no. 1 (2019): 1–57.

From events to data: Politics and the production of government records

Scott J. Cook (Texas A&M University) and David Fortunato (University of California, San Diego)

Issues of measurement have received considerable attention in political science. This is because, as Munck, Møller and Skaaning (2020, 347) argue, “The social sciences. . . are factual sciences, given that they refer to facts about the concrete world. Thus, empirics, and more narrowly measurement. . . are essential parts of social science research.” Many discussions of measurement tend to, rightly, focus on the variety of challenges researchers face in producing data (e.g., concept formation, indicator development, etc.). Here we set aside most of these concerns and focus on a narrower issue: measurement error in government data. Government data—records collected and disseminated by governments and government-adjacent international organizations—are widely used in political science research, including election results, crime statistics, trade data, etc. With these data, researchers evade many of the traditional concerns of measurement, as there tends to be broad consensus on the underlying concepts, indicators are based on facts, and the data are already collected. Despite this, we consistently find that even “high quality” government data are often plagued by mismeasurement, which left unaddressed threatens the validity of not only inferences but basic description.

While political scientists, notably Hollyer, Rosendorff and Vreeland (2014), have studied governments’ choices to report data to external actors, many continue to (implicitly) assume that government data are accurate, or, that the manipulation of government data is exclusive to autocracies. In the following, we discuss how government data collection and reporting processes create opportunities for systematic errors even in wealthy democracies. We then show that misreporting and underreporting is prevalent in official U.S. data on crime and policing—data collected in a well-resourced democracy, purporting to convey objective facts on highly salient issues—which makes obtaining accurate estimates on events like killings by police extraordinarily challenging. This, we argue, is not because the collection of these data is inherently difficult, but instead a logical result of the political processes that shape whether and how government data are produced.

As such, we feel there is a need for further research into the political economy of government data. Scholars must consider the literal data-generating (or data-production) process separately from the outcome-generating process which tends to be the focus of our theoretical interest. Not only are questions on the data-generating process—government transparency, accountability, and the politics thereof—interesting in and of themselves, but they also have clear implications for any social science research utilizing these data. To aid in this, we outline how researchers can begin to engage questions of data production and discuss the consequences of failing to do so.

Politics and the production of government data

Decisions about the collection and dissemination of data by governments *are* policy choices. As such, they warrant scrutiny by researchers, especially political scientists. To focus our discussion, we concentrate on event data—Schrodt (2012, p. 548) defines an event as a “discrete incident that can be located at a single time (usually precise to a day) and set of actors,”—such as a death, vote, payment, etc. The outcome of interest is the occurrence of the event itself, with the determinants of this informing the *event-generating* process. Conditional on the realization of this event, it is either accurately recorded in government data or not, which constitutes the *data-generating* process. Despite our focus on event data here, the concerns we raise on data quality apply broadly to various other types of government data.

From events to data

While the specific data-generating process for any issue is unique, there are a set of minimal questions that researchers should consider when using government data. First, what is the process through which events are memorialized? Are the events automatically submitted into an archive (e.g., weather indicators, legislative proposals), or, do the events require a party to voluntarily report or document them (e.g., crime, death). Second, who enters the event into the record? Reporting the event may be labor or expertise intensive, creating inequities due to differing resources,

From events to data (cont.)

Cook and Fortunato

or, recorders may manipulate the data due to personal political preferences. Third, why was the event recorded, or, for what reason are the data being collected? Some government data, such as those provided by the National Weather Service, are collected for their own sake, in an effort to transparently provide a public good (supporting commerce and public safety). Yet, other data are collected with clear policy objectives in mind, that is, to accomplish a specific task. For example, data on the number of public school students with special needs, the services provided to them, and the performance of those students are required by the No Child Left Behind Act in order for schools to access federal funds provided by the Individuals with Disabilities Education Act. It seems likely that data collected for their own sake as a public good and data collected in pursuit of funding transfers are prone to different types of error in recording, aggregation, and dissemination. Often it will be useful to map out the process by which events become data.

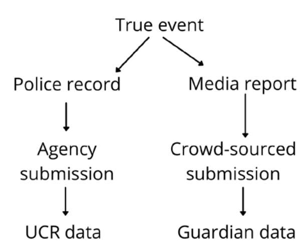


Figure 1: Alternative data-production processes from a single event.

Drawing from recent research by Cook and Fortunato (2022) on killings by police, Figure 1 demonstrates two ways by which a true event may become data: government data, (the FBI's Uniform Crime Reporting [UCR] program) vs. crowd-sourced media reports (The Guardian's The Counted data). Focusing on the UCR data production, we observe that initial reports are filed by individual officers. These reports are aggregated by the agency and submitted to the FBI for inclusion in the UCR database. As such, there are several steps in this process which may cause true events to go unreported in the final UCR data. Individual officers may be tempted to file inaccurate reports (e.g., Egel, Chabria and Garrison, 2017) or not report at all. Individual agencies, then, have discretion on whether to enter events into their UCR submission (or submit to UCR at all). Our research shows that many choose to redact police killings from their reports—for example, the Guardian's The Counted Data verified 71 killings by police in Florida in 2015, yet none are entered into the UCR. Finally, once the data find their way to the UCR, those running the program have additional discretion. For example, the 2016 UCR data, the first released under the Trump cabinet, were conspicuously less detailed than in previous years (Malone and Asher, 2017).

Setting aside incentives for outright manipulation, there is substantial variation in law enforcement agencies' resources for aggregating and submitting data to the UCR; many agencies are quite small, employing few officers and civilian support staff. Between 1995 and 2017, over 25% of agencies had less than three *total* employees, meaning that many police agencies may simply lack the resources to comply with data requests. To illustrate this point, we gather data on UCR participation for all 19,095 state, county, and city police agencies for the 1960-1994 and 1995-2017 periods[1] and regress participation (whether an agency submitted data) on two proxies for agency resources: the total number of agency employees and the size of the population it serves (both rescaled to standard normal). The results in Table 1 indicate large, positive correlations between our proxies for agency resources and UCR participation. This suggests that we likely have significantly less (and poorer) data on crime in smaller (and poorer) communities.

Table 1: UCR compliance and proxies for agency resources

	1960-1994		1995-2017	
Employees	0.026***		0.016***	
	(0.001)		(0.001)	
Population		0.041***		0.034***
		(0.001)		(0.001)
State FE	x	x	x	x
Year FE	x	x	x	x
Observations	668,325	668,325	452,226	452,226
R^2	0.213	0.217	0.115	0.120

Note: *p<0.1 **p<0.05 ***p<0.01

These are critical considerations for applied researchers when using these data. Our reading of the extant literature on crime and policing, however, suggests that most prior research using UCR data has not carefully considered these issues. This is particularly troubling given

1 We separate the periods because they come from different sources. The latter are supplied by the FBI in a standard spreadsheet, the former were parsed from oddly formatted or unformatted text files received as part of our FOIA request.

From events to data (cont.)

Cook and Fortunato

that UCR reports missing data cells as *zeros*. This means that UCR zero counts may indicate that the true number of events was zero, that the agency incorrectly reported a zero, or that the agency failed to report any information on that event-type. Yet, the prevailing method for handling UCR zeros in economics research is to treat them as *true counts*, unless the agency submitted nothing at all that year (e.g., Mello, 2019; Weisburst, 2019). This means that agencies that report the number of officers employed, but no other data, will enter empirical analyses as having zero police killings, zero sexual assaults, zero property crimes, etc.

Consequences of imperfect data

When researchers do not separately consider the outcome- and data-generating processes—accepting the data as (near) perfect memorialization of events—they are of course more likely to draw incorrect inferences. The particular nature of these threats to inference depends on how these data are used and whether the outcome- and data-generating process share common determinants. For example, if the events themselves are the unit-of-analysis, then unreported events would induce a form of sample selection bias. More typically, events are located and aggregated into spatial-temporal units (e.g., state-year), where unreported events instead induce measurement error in the outcome. At best, this will produce attenuation toward the null, however, we cannot safely assume this as many of the same features that cause the outcome also cause variation in reporting rates, risking bias in either direction (Carroll et al., 2006)[2]. For example, Glaeser and Sacerdote (1999) use data from the UCR and the National Crime Victimization Survey (NCVS) to compare crime incidence across municipalities. They find a very large, positive correlation between population and crime (more crimes per resident as population grows), but puzzle over the substantially smaller correlation between crime and population—a correlation that is *negative* when comparing cities of 25,000 and greater—when examining survey responses. Given the UCR reports missing values as zeros and the strong correlation between city size (and agency staffing) and agencies' propensity to report data into the UCR, the more likely reason for this gap is that the survey data (assuming the sample is well-calibrated and representative) are providing a more accurate estimate of crime victimization.

Assuming for the moment that the NCVS provide accurate estimates of crime victimization, what would have to be true in order for the UCR and NCVS to provide effectively identical estimates? Returning to discussion above, 1) all victims must report their victimization; 2) all responding officers must accurately memorialize the incident and file the report (a step Eckhouse, 2021 demonstrates is prone to significant manipulation); 3) the agency must submit all reported crimes into the UCR; 4) the FBI releases all UCR data to the public. Without considering agencies' incentives to manipulate, this process chain allows (at least) four opportunities for attrition—victims may choose not to report, responding officers may make filing errors, agencies may fail to comply with UCR, the FBI may not release all information—but almost no opportunities for over-counts apart for a small number false-reports (which are themselves a crime) in step 1. That is, even if the only error in the process is "random," the net effect is still inherently asymmetric, producing lower estimates of the base rate of crime. Given the relationship between agency resources and reporting, these errors are more likely (in practice, larger undercounts of crime) in smaller or poorer cities, inducing further bias. These errors have implications not only for academic research, but government policy, as policymakers utilize these data unaware of their limitations.

Conclusion

Our aim is to present potential issues in the collection, aggregation, and dissemination of government data. While these data are widely used to study many events of interest (e.g., auto accidents, high school graduation rates, unemployment, etc), too often researchers fail to consider their limitations. All government data have a formally or informally mandated data-generating process that risks error given the nature of the process, the actors involved, and their incentive structure. We illustrate opportunities for significant manipulation and selection bias using the data-generating process underlying the UCR that there is strong positive correlation between UCR compliance and agency resources, and discussing researchers' insufficient consideration of the UCR's data-generating process. Because of this failure, it is our judgement that nearly

2 This problem is further compounded if both the outcome and input data come from the same source, risking potential "common source" bias (Favero and Bullock, 2015).

From events to data(cont.)

Cook and Fortunato

every study of crime or policing which naïvely employs UCR data cannot be trusted. How many other fields of study treat similarly imperfect data as unbiased samples?

While we have focused on data from government records, these issues broadly apply to other data sources (e.g., media reports, historical accounts) frequently used in political science research. In peace studies, for example, researchers often use data on political violence (ex. Social Conflict in Africa Database) drawn from news media reports (ex. the *Associated Press*). As a result, events in some countries (e.g., lesser developed states) tend to be underreported (Hendrix and Salehyan, 2015), and, even within a particular country, events in some areas (e.g., capital cities) are more likely to be reported (Weidmann, 2016). The systematic errors in the reporting of these events have consequences identical to those above (as discussed in Cook and Weidmann, 2019). In light of this, many conflict researchers have increasingly used official data (Weidmann, 2015) or conflict archives (Balcells and Sullivan, 2018) when available. In some cases these non-media data may be more reliable, however, it should not be assumed that they will be given our discussion here. Instead, we encourage researchers to scrutinize their data—asking at least how events are recorded, who records them, and for what purpose—regardless of the original source(s).

Beyond the simple recognition of the limitations in one's data, how should researchers proceed? While much of the specifics will have to be addressed elsewhere, in short there are two ways forward. First, where possible, researchers should try to find multiple sources of data on their phenomena of interest, ideally ones in which the preferences and the priorities of the data producers diverge (e.g., contrasting data collected by police departments with data collected by media). Minimally, this will allow researchers to compare findings across alternative data and ensure that any inferences drawn are not source, and therefore *process*, sensitive.

Second, analysts should consider explicitly modeling their uncertainty over data quality. Without additional data, this will typically take the form of sensitivity analysis or bounding, approaches that may be particularly helpful when the shape of the potential bias can be inferred (e.g., Knox, Lowe and Mummolo, 2020). With multiple sources of data, the available options are much richer, as most measurement error models require some type of validation or replication data (Carroll et al., 2006). For example, Cook et al. (2017) demonstrate how with two sources of conflict data, researchers can analyze both the probability of event and report, that is, specify models of both the outcome- and data-generating processes, respectively. We feel that future research in this area is especially worthwhile, as the variety of available data sources and types continues to grow. As such, better understanding how to effectively integrate multiple sources of data to obtain more accurate results is likely to be a fruitful area for research.

References

Balcells, Laia and Christopher M Sullivan. 2018. "New findings from conflict archives: An introduction and methodological framework."

Carroll, Raymond J, David Ruppert, Leonard A Stefanski and Ciprian M Crainiceanu. 2006. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC.

Cook, Scott J, Betsabe Blas, Raymond J Carroll and Samiran Sinha. 2017. "Two wrongs make a right: Addressing underreporting in binary data from multiple sources." *Political Analysis* 25(2):223–240.

Cook, Scott J. and David. Fortunato. 2022. "The Politics of Police Data: State Legislative Capacity and the Transparency of State and Substate Agencies." Working Paper.

Cook, Scott J and Nils B Weidmann. 2019. "Lost in Aggregation: Improving Event Analysis with Report-Level Data." A*merican Journal of Political Science* 63(1):250–264.

Eckhouse, Laurel. 2021. "Metrics Management and Bureaucratic Accountability: Evidence from Policing." *American Journal of Political Science* in press.

Egel, Benjy, Anita Chabria and Ellen Garrison. 2017. "Hands removed, findings changed: Pathol- ogists say San Joaquin sheriff 'does whatever he feels like doing'." *Sacramento Bee* December 05, 2017 .
URL: *Access*

From events to data (cont.)

Cook and Fortunato

Favero, Nathan and Justin B Bullock. 2015. "How (not) to solve the problem: An evaluation of scholarly responses to common source bias." *Journal of Public Administration Research and Theory* 25(1):285–308.

Glaeser, Edward L and Bruce Sacerdote. 1999. "Why is there more crime in cities?" *Journal of political economy* 107(S6):S225–S258.

Hendrix, Cullen S and Idean Salehyan. 2015. "No news is good news: Mark and recapture for event data when reporting probabilities are less than one." *International Interactions* 41(2):392–406.

Hollyer, James R, B Peter Rosendorff and James Raymond Vreeland. 2014. "Measuring transparency." *Political analysis* 22(4):413–434.

Knox, Dean, Will Lowe and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* pp. 1–19.

Malone, Clare and Jeff Asher. 2017. "The First FBI Crime Report Issued Under Trump Is Missing A Ton Of Info." *FiveThirtyEight* October 27, 2017. URL: Access

Mello, Steven. 2019. "More COPS, less crime." *Journal of Public Economics* 172:174–200.

Munck, Gerardo L., Jørgen Møller and Svend-Erik Skaaning. 2020. Conceptualization and Mea surement: Basic Distinctions and Guidelines. In The *SAGE Handbook of Research Methods in Political Science and International Relations*, ed. Luigi Curini and Robert Franzese. London: SAGE Publications chapter 19, pp. 331–352.

Schrodt, Philip A. 2012. "Precedents, Progress, and Prospects in Political Event Data." *International Interactions* 38(4):546–569. URL: *https://doi.org/10.1080/03050629.2012.697430*

Weidmann, Nils B. 2015. "On the accuracy of media-based conflict event data." *Journal of Conflict Resolution* 59(6):1129–1149.

Weidmann, Nils B. 2016. "A closer look at reporting bias in conflict event data." *American Journal of Political Science* 60(1):206–218.

Weisburst, Emily K. 2019. "Safety in police numbers: Evidence of police effectiveness from federal cops grant applications." *American Law and Economics Review* 21(1):81–109.

Moving Beyond Newspapers and Text to Study Contentious Politics[1]

Zachary C. Steinert-Threlkeld (University of California, Los Angeles)

THE PROBLEM

Event datasets should use images, especially ones from social media, to generate their data. Event data are data where each observation consists of, minimally, an actor performing an action directed at a second actor at a location on a date. At its most expansive, the action can be any type of directed political action, such as voting on a bill or giving a speech. In the context of contentious politics, the field of focus for this essay, commonly used datasets with global coverage are the Armed Conflict Locations and Event Data (ACLED) (Raleigh et al. 2010), Integrated Conflict Early Warning System (ICEWS) (Boschee et al. 2015), Nonviolent and Violence Campaign Outcomes (NAVCO) 3.0 (Chenoweth, Pinckney, and Lewis 2018), and Uppsala Conflict Data Program – Georeferenced Event Dataset (UCDP-GED) (Sundberg and Melander 2013). There are also dozens of event datasets of narrower geographic or temporal scope.

Event datasets vary along two dimensions, the variety of raw data used and the type of processing applied to those data. Raw data usually takes the form of newspaper articles, though some research makes use of archives or non-governmental organization reports (Davenport and Ball 2002; Sullivan 2016). Since almost all datasets use one type of raw data, this first dimension varies based on the breadth of raw data ingested. For example, the Dynamics of Collective Action in the United States project uses one newspaper, *The New York Times*, while ICEWS ingests several hundred (McAdam et al. 2009). The type of processing takes one of three forms: fully manual (humans read items to extract events), fully automatic (a computer uses a dictionary or is trained on labeled data to identify events from raw data), or hybrid (a computer filters large quantities of raw data and humans code the filtered data). Fully automatic datasets tend to use the Conflict and Mediation Event Observations (CAMEO) event ontology of 20 ordinal event types ranging from public statements to unconventional mass violence (Gerner et al. 2002).

In theory, a third dimension, the type of raw data (text, audio, image, or video) exists, though in practice only text has been used. Event datasets' exclusive focus on text has been because newspapers produce more text than images and, for automatic approaches, computers were until recently not proficient at interpreting complex images.

Regardless of data variety or processing type, most existing event data are constrained by the same problem: they rely on newspapers, and newspapers suffer large, known, replicated biases (Donnay and Filimonov 2014; Myers and Caniglia 2004; Weidmann 2014). These biases arise because newspapers seek to maximize profits (circulation) within production cost constraints. Costs impose space constraints, forcing publishers to select stories for their ability to maximize purchases *regardless of how representative the chosen events are of actual events*. Because humans are attracted to negative information (conflict) and novelty, production constraints cause newspapers to emphasize large and unexpected events (Baumeister, Bratslavsky, and Vohs 2001). New methods, no matter how advanced, can surmount the intrinsic biases in the newspaper data generating process.

In the context of protests, these structural features mean event datasets based on newspapers over-report large and violent protests. Small, peaceful, or long-duration ones are less likely to receive coverage and enter event datasets; for a more thorough explanation of how the data generating process introduces bias, see Steinhardt and Goebel (2019). Even if newspapers did not face bias-causing incentives, their limited space would mean they could not publish information on every protest. In fact, this limited production space means the use of alternate data sources such as social media records orders of magnitude more true positive protests, primarily due to the recording of smaller events (Zhang and Pan 2019). Finally, the increasing precariousness of newspapers means they cannot report on as much activity as in decades past, so ongoing reliance on them will generate fewer observations (Peterson 2021). If declining circulation causes newspapers to publish more sensational events, it will also increase the bias in event data.

MOVING BEYOND NEWSPAPERS

The most common solution for addressing pub-

1 Karsten Donnay and Han Zhang provided valuable feedback on an early version of this essay.

Moving Beyond Newspapers (cont.)

Steinert-Threlkeld

lication bias has been to source more newspapers. Though admirable, this approach relies on a research team creating a list of publications to ingest, and the exhaustiveness of the resulting publication list will always be limited by the size of the team and the breadth of their expertise. Local publications are also less likely to be online and more likely to not be in English, which introduces translation complications.

The best way to push the variety dimension is to crowdsource event reports: when anyone can submit evidence of an event, the size of the research team is no longer limited by money available to researchers to hire assistants or time free to train them or computer algorithms. When anyone in the world is allowed to be a source of raw data, the amount of raw data available to researchers increases dramatically. The Crowd Counting Consortium (CCC) and Cornell Labor Relations Tracker (CLRT) are two recent event datasets that demonstrate the power of crowdsourced data. The CCC started in 2017 to document that year's Women's March, and it is actively maintained by a small set of researchers. Its homepage provides not only data but also a submission form, and this submission form is how it identifies events. Submissions are verified before being incorporated into the dataset, and submitters must provide documentation of the event. A plurality of the events derive from newspapers, but over half are from television or social media posts; for the 2017 Women's March, 26.25% of events were from television stories, 26.84% were from Twitter or Facebook posts, and 46.91% were from newspapers (Sobolev et al. 2020). The CLRT started in late 2020 and documents strikes and labor protests. In addition to ingesting traditional sources, it identifies events from social media posts, reddit, and even personal exchange with organizers and allows for direct submission of events. The CLRT has not publicly released its data, but studies of labor contention are also known to underreport events (see Robertson (2007) for a discussion of these difficulties), so it is reasonable to assume that the dataset is significantly more expansive than predecessors.

ACLED, the most comprehensive actively maintained contentious politics dataset, uses a soft version of crowdsourcing. The overall project relies heavily on the active monitoring of local and regional news sources, but it also uses local trusted individuals to provide information about events. This reliance on local informants is especially important in countries with restricted media. In Ethiopia, for example, such sources have provided 15% of events, and methodology documents ACLED provides for other countries reveals similarly diverse sourcing[1]. This crowdsourcing is "soft" because ACLED's curates its crowd: it relies on social media posts or reports from trusted sources and is necessarily guarded about the extent and identity of them. Nonetheless, it is clear that a substantial percentage of events ACLED records would not have appeared if the research team relied only on newspapers.

Social media is the easiest source of crowd sourcing once the production of content is reconceptualized as the submission of potential events. The next frontier for event data is therefore to record events automatically from social media posts. This approach is promising because the size of the crowd is only limited by the size of the social media platform; the number of potential sources number in the millions. Automating collection and processing allows the resulting dataset to be more expansive than the single-country focus of the CCC and CLRT, a result of their manual review of posts. ACLED is not automatic but is global because it is highly resourced. Its capabilities are therefore out of reach of small teams of researchers, which is to say the vast majority of us.

Potential sources of social media data include Facebook, Instagram, reddit, Twitter, VKontakte, Sina Weibo, YouTube, and chat applications like WhatsApp and Telegram. Since most Facebook accounts are private, it is not a useful source of data; CrowdTangle, which provides data on Facebook Pages, is similarly not amenable to automation because researchers must know which Pages' data they want. Instagram's API is very restrictive, though Python packages exist that facilitate scraping. Reddit is promising, though its posts tend to redirect one to other platforms, increasing programming work significantly. Twitter has a norm of public accounts and a well-documented API that can provide millions of posts per day, but its reach is not as large as Facebook or Instagram and version 2 of the API will cause smaller events to go unreported compared to the API it replaces. VKontakte is easily accessible and widely used in Slavic countries, but English language documentation and support is scarce. Sina Weibo, China's Twitter, is widely used but heavily censored, and two citizens who made a dataset of protests using its posts (Wickedonna) were

1 https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Ethiopia-Sourcing-Profile_February-2020.pdf

Moving Beyond Newspapers (cont.)

Steinert-Threlkeld

jailed in 2016 and are now under constant surveillance (Wong 2020). YouTube's API is generous, but location information, necessary for event data, is close to non-existent. Chat applications require ongoing researcher involvement and are therefore not amenable to automatic data collection.

Despite the limitations of Twitter, it is the most promising source of event data because of its global reach, extensive documentation, and norm of public production with greater than appreciated amounts of geolocation (Steinert-Threlkeld 2018). Though no event datasets based on Twitter exist that are on the scale of ICEWS, ACLED, or NAVCO 3.0 , there is growing use of it for the study of contentious politics. See, for example, its use in Kenya (Dowd et al. 2020); the study of election violence in Ghana, the Philippines, and Venezuela (Muchlinski et al. 2020); and state and protester violence affected protests in Chile, Hong Kong, Spain, South Korea, and Venezuela (Steinert-Threlkeld, Chan, and Joo 2022). These approaches rely on the streaming endpoint of Twitter's API and subsequent processing of the returned tweets to identify events. Identification can be as simple as a dictionary of relevant keywords but usually involves using a machine learning classifier trained on labeled data, plus subsequent processing of this smaller set of data to generate an event dataset that is analyzed for the particular research project.

Relying on Twitter to generate event data leads to two primary concerns, duplication and bias. With automatic processing, the risk of duplicate events being coded as separate events is higher than for manual or hybrid approaches, and the profusion of duplicate events in GDELT is one of the reasons that dataset is not commonly used in political science (Caren 2014). This concern is valid and addressable. For example, retweets can be filtered out so that they do not create multiple events; tweets can be aggregated to place-day; or only the first tweet mentioning an event is kept. As to bias, Twitter users in the United States are not a representative sample of Americans (Smith, Anderson, and Caiazza 2018) or Women's March protesters (Barrie and Frey 2021), and this composition bias probably exists in other countries[2]. However, a biased sample is only a problem if its causes biased behavior in the researcher's domain. For many researchers, behavior of a biased sample (Twitter) probably does not differ from behavior of a representative sample (the "real world"). For example, protesters in France are more closely connected to each other in their Twitter network (Larson et al. 2019), just like in offline networks in Hong Kong (Bursztyn et al. 2021). More broadly, social network structures appear similar online and off (Bisbee and Larson 2017).

In addition, the concern about social media bias forgets a key fact: newspapers are biased. The use of social media data should therefore not be predicated on whether they are unbiased but whether they are less biased than newspapers. Since social media companies do not face constraints that force them to limit the quantity of content published on their platforms and the vast majority of users publish from intrinsic motivation (Marwick and Boyd 2010), the representation of the world retrieved from social media content should be more accurate than that from newspapers. Comfort with very biased sources (newspapers) should not prevent the embrace of new ones that, for clear theoretical reasons, almost certainly exhibit less bias (Steinhardt and Goebel 2019).

2 Bias may not exist in some countries where Twitter has its greatest market penetration, e.g. like the Gulf States, Indonesia, or Japan.

MOVING BEYOND TEXT

Another limitation of existing event datasets is the briefly mentioned third dimension: almost all event datasets are based on text. Text is easily processed but limited in the types of variables it can record. For example, police and protester violence is always operationalized as a nominal or ordinal variable. This coarsening limits the variation in event datasets, which in turn impedes the study of subnational violence. Images are more subtle than text and can therefore measure behavior better (Joo and Steinert-Threlkeld Forthcoming).

Twitter, and social media more broadly, is particularly promising for the study of event data because they frequently contain images; in my experience, 10% of tweets contain an image. Images are important because they can measure difficult to operationalize variables of theoretical interest to various strands of research (Z. C. Steinert-Threlkeld 2019). For example, images can reveal emotional states of protesters, an important factor in collective action that has resisted observational operationalization (Pearlman 2013; Won, Steinert-Threlkeld, and Joo 2017); protest size, one of the strongest correlates of protest success (Sobolev et al. 2020), and protester sex, which many theorize affects mobilization (Barrie and Frey 2021; Schaftenaar 2017; Urdal and Hoelscher 2012). It is straightforward to deduplicate images, and therefore

Moving Beyond Newspapers (cont.)

Steinert-Threlkeld

events, using the final fully-connected layer of a convolutional neural network (the classifier); while the bias concern remains, those who share images on Twitter do not appear to differ from those who do not.

As social scientists grow comfortable incorporating computer vision techniques into their work, the use of images to study contentious politics has increased. Mitts, Phillips, and Walter (2021) shows that expressions of support for terrorist groups is affected by the types of videos they broadcast. Zhang and Pan (2019) generates the most comprehensive dataset on protests in China using text and images shared on Sina Weibo; they also show that social media recover orders of magnitude more events than traditional news sources. Steinert-Threlkeld, Chan, and Joo (2022) uses protest images shared in geolocated tweets to argue that state repression has an n-shaped effect on subsequent protests. It also recovers many more protests than newspaper-based event datasets, suggesting the less bias found in Zhang and Pan (2019) exists across media environments.

CONCLUSION

The systematic production of event data has been a core part of the study of contentious politics since the start of the Cold War, and the production of datasets has made important leaps in conjunction with the advance of computers. The field, however, has reached the limit of this approach, especially since the decline of newspapers means there are fewer absolute recorded events to parse. Future development effort should therefore prioritize data sources that are not newspapers. Social media are particularly promising, not least because of the extensive dissemination of images. Measured in file size, a picture is not worth 1,000 words; it is actually worth around 100,000 words, a book. The time is ripe for a new chapter.

REFERENCES

Barrie, Christopher, and Arun Frey. 2021. "Faces in the Crowd: Twitter as Alternative to Protest Surveys." Plos One 16(11): e0259972. http://dx.doi.org/10.1371/journal.pone.0259972.

Baumeister, Roy F, Ellen Bratslavsky, and Kathleen D Vohs. 2001. "Bad Is Stronger Than Good." Review of General Psychology 5(4): 323–70.

Bisbee, James, and Jennifer M. Larson. 2017. "Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity." American Political Science Review 111(3): 502–21.

Boschee, Elizabeth et al. 2015. "ICEWS Coded Event Data." http://dx.doi.org/10.7910/DVN/28075.

Bursztyn, Leonardo et al. 2021. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." American Economic Review: Insights 3(2): 233–50.

Caren, Neal. 2014. "It Is Time to Get Rid of the E in GDELT." Bad Hessian. http://badhessian.org/2014/05/it-is-time-to-get-rid-of-the-e-in-gdelt/.

Chenoweth, Erica, Jonathan Pinckney, and Orion Lewis. 2018. "Days of Rage : Introducing the NAVCO 3.0 Dataset." Journal of Peace Research 55(4): 524–34.

Davenport, Christian, and Patrick Ball. 2002. "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995." Journal of Conflict Resolution 46(3): 427–50.

Donnay, Karsten, and Vladimir Filimonov. 2014. "Views to a War: Systematic Differences in Media and Military Reporting of the War in Iraq." EPJ Data Science 3(1): 1–29.

Dowd, Caitriona, Patricia Justino, Roudabeh Kishi, and Gauthier Marchais. 2020. "Comparing 'New' and 'Old' Media for Violence Monitoring and Crisis Response in Kenya." Research and Politics 7(3): 1–9.

Earl, Jennifer, Andrew Martin, John D. Mccarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." Annual Review of Sociology 30: 65–80.

Gerner, Deborah J., Philip A. Schrodt, Rajaa Abu-Jabr, and Omur Yilmaz. 2002. "Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions." In Annual Meeting of the International Studies Association,.

Joo, Jungseock, and Zachary C. Steinert-Threlkeld. 2022. "Image as Data: Automated Content Analysis for Visual Presentations of Political Actors and Events." Computational Communication Research.
Larson, Jennifer M, Jonathan Nagler, Jonathan Ronen, and Joshua A Tucker. 2019. "Social Networks and Protest Participation: Evidence from 130 Million Twitter Users." American Journal of Political Science 63(3): 690–705.

Moving Beyond Newspapers (cont.)

Steinert-Threlkeld

Marwick, a. E., and D. Boyd. 2010. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." New Media & Society 13(1): 114–33. http://nms.sagepub.com/cgi/doi/10.1177/1461444810365313 (July 10, 2014).

McAdam, Doug, John McCarthy, Susan Olzak, and Sarah Soule. 2009. "Dynamics of Collective Action Dataset." https://web.stanford.edu/group/collective-action/cgi-bin/drupal/.

Mitts, Tamar, Gregoire Phillips, and Barbara F. Walter. "Studying the Impact of ISIS Propaganda Campaigns." Journal of Politics: 1–10. https://www.journals.uchicago.edu/doi/pdf/10.1086/716281.

Muchlinski, David et al. 2020. "We Need to Go Deeper: Measuring Electoral Violence Using Convolutional Neural Networks and Social Media." Political Science Research and Methods: 1–18.

Myers, Daniel J., and Beth Schaefer Caniglia. 2004. "All the Rioting That's Fit to Print: Selection Effects in National Newspaper Coverage of Civil Disorders, 1968-1969." American Sociological Review 69(4): 519–43.

Pearlman, Wendy. 2013. "Emotions and the Microfoundations of the Arab Uprisings." Perspectives on Politics 11(02): 387–409.

Peterson, Erik. 2021. "Paper Cuts: How Reporting Resources Affect Political News Coverage." American Journal of Political Science 65(2): 443–59.

Raleigh, Clionadh, Andrew Linke, Havard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." Journal of Peace Research 47(5): 651–60.

Robertson, Graeme B. 2007. "Strikes and Labor Organization in Hybrid Regimes." American Political Science Review 101(04): 781–98.

Schaftenaar, Susanne. 2017. "How (Wo)Men Rebel: Exploring the Effect of Gender Equality on Nonviolent and Armed Conflict Onset." Journal of Peace Research 54(6): 762–76.

Smith, Aaron, Monkia Anderson, and Tom Caiazza. 2018. Social Media Use in 2018.

Sobolev, Anton, Jungseock Joo, Keith Chen, and Zachary C. Steinert-Threlkeld. 2020. "News and Geolocated Social Media Accurately Measure Protest Size Variation." American Political Science Review: 1–9.

Steinert-Threlkeld, Zachary C. 2018. Twitter as Data. Cambridge University Press.

———. 2019. "Comment: The Future of Event Data Is Images." Sociological Methodology 49.

Steinert-Threlkeld, Zachary, Alexander Chan, and Jungseock Joo. 2022. "How State and Protester Violence Affect Protest Dynamics." The Journal of Politics.

Steinhardt, H. Christoph, and Christian Goebel. 2019. Better Coverage, Less Bias: Using Social Media to Measure Protest in Authoritarian Regimes. https://www.researchgate.net/publication/332112415_Better_coverage_less_bias_Using_social_media_to_measure_protest_in_authoritarian_regimes.

Sullivan, Christopher M. 2016. "Political Repression and the Destruction of Dissident Organizations." World Politics 68(4): 645–76.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." Journal of Peace Research 50(4): 523–32.

Urdal, Henrik, and Kristian Hoelscher. 2012. "Explaining Urban Social Disorder and Violence: An Empirical Study of Event Data from Asian and Sub-Saharan African Cities." International Interactions 38(4): 512–28.

Weidmann, N. B. 2014. "On the Accuracy of Media-Based Conflict Event Data." Journal of Conflict Resolution 59(6): 1129–49.

Won, Donghyeon, Zachary C. Steinert-Threlkeld, and Jungseock Joo. 2017. "Protest Activity Detection and Perceived Violence Estimation from Social Media Images." In Proceedings of the 25th ACM International Conference on Multimedia, , 786–94.

Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." Sociological Methodology 49: 1–48.