

The Challenges of Organizational Classification: A Research Note

Social Currents

1–8

© The Southern Sociological Society 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2329496519878469

journals.sagepub.com/home/scu

Samantha Plummer¹ , Melanie M. Hughes¹ ,
and Jackie Smith¹

Abstract

Social movement, civil society, and world polity scholars use counts of nongovernmental organizations (NGOs) to evaluate important theoretical and empirical claims. To construct these measures, researchers often classify NGOs by their goals and/or domains. However, over time, the ways organizations describe and orient themselves change, blurring boundaries between organizations and complicating measurement. In this research note, we identify methodological challenges of organizational classification in the context of our work constructing a longitudinal dataset of transnational social movement organizations. We draw attention to an understudied cause of measurement error: overcounting of organizations. We suggest that as automated methods for classifying data become widespread, devising strategies for dealing with these challenges becomes even more pressing.

Keywords

organizations, classification, social movements, NGOs, methodology

Introduction

Civil society, the space of collective action and voluntary association, is a central concept in political science, sociology, and organizational studies; it is also notoriously difficult to measure empirically. To assess civil society's strength and reach, scholars often use counts of nongovernmental organizations (NGOs), a highly diverse set of professional, scientific, cultural, and political associations (for a general review, see Meyer 2010). Whereas some civil society theories are operationalized through counts of entire populations of NGOs, others theorize the activities or impacts of particular kinds of associations and therefore measure subsets of organizations. Researchers who classify NGOs by their *goals* distinguish

organizations that, for example, seek social change—advocacy, activist, or social movement organizations—from the broader population of professional and cultural associations (e.g., Hughes et al. 2015; Plummer, Smith, and Hughes 2018; Smith and Wiest 2012). Researchers who classify NGOs by their *domains* identify organizations working in a particular focus area or sphere, such as development (e.g., Murdie and Davis 2012), education (e.g., Kim and Boyle 2012), the environment

¹University of Pittsburgh, PA, USA

Corresponding Author:

Samantha Plummer, Department of Sociology, University of Pittsburgh, 2400 Posvar Hall, 230 Bouquet St., Pittsburgh, PA 15260, USA.

Email: Ssp19@pitt.edu

(e.g., Andrews and Edwards 2005), labor (e.g., Martin et al. 2006), and women's and human rights (e.g., Hughes et al. 2018). Although the enumeration of goals and domains of NGOs has yielded important insights into the structure and influence of civil society, methodological attention to how researchers classify organizations into these goals and domains is sparse.

The extant methodological literature on counting civil society organizations has focused principally on assessing the representativeness and comprehensiveness of popular data sources. Scholars are concerned that organizational directories systematically fail to identify groups that meet criteria for inclusion, distorting our understanding of organizations and their activities (Andrews et al. 2016; Bevan et al. 2013; Brulle et al. 2007; Edwards and Foley 2003; Grønberg 1994; Martin et al. 2006; Reith et al. 2016). Researchers have paid considerably less attention to the inverse possibility of mistaken organizational inclusion—where organizations are classified as having goals or working in domains when they do not. Such overcounting is problematic for research on the structure, behavior, and impact of organizations.

The need for methodological attention to (mis)classification has grown as a result of interconnected changes in the strategic behavior of organizations and in the social environments where organizations operate. As isomorphic pressures¹ have grown in intensity, the boundaries between social movement, civil society, state, and corporate actors have become increasingly blurry (Bromley and Meyer 2017; de Bakker et al. 2013; Mooney 2012; Smith et al. 2017). Organizations with different goals have adopted similar structures, discourses, and practices, and new, hybrid forms of organization that resist easy classification have emerged (Hasenfeld and Gidron 2005). Social movements' successes in agenda setting have led to the appropriation of their issue frames by nonactivist organizations and corporations (Goldman 2005; Pearce 2013). And, social movements' networking and coalition building has fueled the formation of

multi-issue groups that are not easily divided into distinct domains (Smith and Wiest 2012).

Attending to these challenges is critically important as data sources are increasingly digitized, enabling the partial or full automation of text classification. Computer-assisted classification dramatically decreases the manual coding required to prepare data for analysis, thereby saving researchers' (and data publishers') time and money. Yet, as Putnam (2016) points out, these techniques unlock "shortcuts that enable ignorance as well as knowledge" (p. 379). If organizations with different goals and working in different domains use similar language to describe themselves, keyword-based techniques will misclassify large numbers of organizations and possibly mistakenly include organizations in their counts and analyses. Supervised machine learning, where computers are trained to classify data in more complex ways, requires that humans properly classify data to train the computer. The accuracy and reliability of a coder, automated or not, rely on the quality of classifications.

In this research note, we discuss the challenges of classifying organizations by goal and domain in the context of our experience extending Smith's dataset, *Transnational Social Movement Organizations 1953–2003*, to 2013 (see Plummer et al. 2018; Smith and Wiest 2012). Our research suggests not only that classification is becoming more difficult but also that researchers should pay more attention to the problem of overcounting; similar to a type 1 error, or "false positive," overcounting occurs when researchers count organizations that do not fit their selection criteria. We suggest that keyword-driven classification is not an appropriate strategy in all occasions. This note should be of interest to a broad readership including social movement, world polity, and organizations researchers who construct datasets and whose work involves counts of groups, as well as computational social scientists who create programs and teach machines to classify data. We begin with a brief description of our research before turning to detailed descriptions of these challenges and our methods for handling them.

Brief Background on the Project

The methodological challenges articulated here arose from our efforts to update the *Transnational Social Movement Organizations (TSMO) Dataset, 1953–2003*, a biennial dataset with information on international nongovernmental groups that advance social or political change (for more details, see Smith and Wiest 2012). We coded information about organizations—including aims, activities, structure, and country membership—from the annual *Yearbook of International Organizations*, the most common source of count data on international nongovernmental organizations (INGOs) (Union of International Associations [UIA] 2005–2013).

Because the *TSMO Dataset* is designed to provide measures of social movement organizations (SMOs), our first task was to decide which organizations had an explicit purpose of promoting political or social change, broadly defined. Then, using information on stated aims and goals, we classified each organization into one or more substantive areas (e.g., women's rights). We discuss methodological challenges we faced at each of these two stages of classification. Because of our research team's substantive interests in environmental and women's rights organizations, the examples we use focus chiefly on these two issue areas.

Classification Challenges

Sources of Classification Issues

Organizational research has been facilitated by the emergence of organizational directories designed to enable organizational networking and coordinating. Such directories include the *Encyclopedia of Organizations*, *Yearbook of International Organizations*, government registries, and domain-specific databases produced by Human Rights Internet and the National Center for Charitable Statistics (NCCS), among others. Researchers have long made use of such directories, but it is not always the case that these directories' criteria for inclusion and categorization, or the rigor with which these criteria

are applied, match those of researchers. Thus, researchers must either devise procedures for identifying records of interest and/or supplement directories with additional sources of information to compile complete accounts of relevant organizational populations.

To categorize organizations by goal (e.g., activism, philanthropy, research, service provision) and domain (e.g., women's rights, environment, democracy), researchers devise their own classification systems or rely on those provided by data publishers. The validity of these systems has arguably been complicated by three processes: (1) discursive isomorphism—the propensity for organizational actors to develop similar ways of thinking and talking about shared problems, (2) coalition building, and (3) multi-issue framing. Figure 1 depicts the relationship between these processes and goal and domain classification issues. We discuss these relationships in turn.

Discursive isomorphism occurs when organizations frame their work in ways that resonate with mainstream ideas and/or attract support from other organizations in the larger environment. Organizations frame their goals and domains in relation to their discursive opportunity structure—the constellation of ideas, vocabularies, and institutions that are accessible and legitimate in a given political culture (Koopmans and Statham 1999). To attain or maintain legitimacy, appeal to targets of mobilization, and/or gain funding, organizations may strategically select issues and frames deemed likely to resonate with mainstream ideas and pass the agenda-vetting process of gatekeepers (Bob 2009; Carpenter 2014; Wong 2012).

Not all organizations desire frame resonance. Because they challenge the status quo, SMOs are more likely than other kinds of organizations to create and adopt nonresonant or "radical" frames. But, if an SMO's issue agenda is adopted by organizations that are highly central to a given advocacy network, the former's discourse is liable to be taken up by a variety of organizations and actors in and outside the network (Carpenter 2014). This process of diffusion can result in the appropriation or cooptation of movement discourse, that

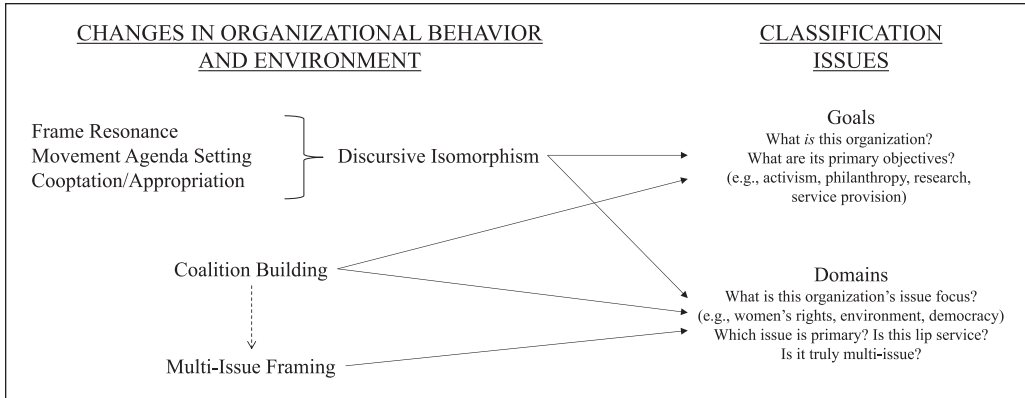


Figure 1. Sources of organizational classification issues.

is, the espousal of its content and marginalization or subversion of its intent (Burke and Bernstein 2014; Smith et al. 2017).

Together, frame resonance, agenda setting, and cooptation/appropriation generate classification problems for researchers seeking to categorize organizations by goal or domain. The mainstreaming of activist discourse like “women’s empowerment” and “sustainable development” to nonactivist organizations increases the difficulty of classifying organizations by goal—if it looks like an SMO and talks like an SMO is it an SMO?—and by domain—if it claims to support sustainable development is it an environmental organization?

Organizations’ formation of coalitions can also generate goal and domain classification issues. When organizations with diverse goals (e.g., companies, governmental agencies, and civil society groups) build a coalition to address a particular issue, it can be complicated to classify that coalition by goal; directory profiles do not always make clear which member organizations’ understandings and strategies for action prevail within a given coalition. Coalition building by organizations that have similar goals but work on different domains (e.g., big tent social movement coalitions) can lead directly to domain classification challenges; it is often difficult to determine which issues such coalitions are actively working on in a given time period.

Coalition building by these organizations can also lead indirectly to domain classification challenges when member organizations develop multi-issue frames in the process of forming coalitions.

As organizations and activists engage in cross-domain formal coalition building and/or informal networking, they become more aware of the connections between different global concerns and generate multi-issue frames that reflect those connections.² When organizations develop multi-issue frames, they are neither creating a common language around the same problems nor necessarily signaling that they work on all the problems they reference. Instead, they are indicating the connections between those problems. Multi-issue framing presents a challenge for researchers who seek to reduce the complexity of organizations’ domains to productively think about and analyze how different organizational populations emerge, die, and change over time. Multi-issue organizations present a classification challenge not only for scholars of international organizations but also for scholars of domestic nonprofits; the NCCS’s National Taxonomy of Exempt Entities (NTEE) codes, “the most commonly accepted categorization of nonprofit organizations in the United States,” to a great extent misrepresent organizations that work in more than one domain (Fyall, Moore, and Gugerty 2018:678).

Examples of Classification Issues

In the course of our research, we encountered several goal and domain classification issues. To identify whether an organization was an SMO, we drew primarily from the “aims” section of the *Yearbook*, which specifies the organization’s issue agenda and typically suggests a political orientation. Searching for new groups, we noticed that issue areas that had once fallen under the near-exclusive purview of SMOs had been adopted by a broader range of NGOs that did not clearly fit our SMO category, complicating our classification technique. For example, the 2013 edition of the *Yearbook* includes the Global Women Petroleum and Energy Club, which, according to its aims statement, was founded in 2000 to “recognise the ascending and significant role of women fulfilling key roles in global oil, gas and energy industry. The Club has provided a platform ever since, always with women’s advancement at its heart.” This organization advocates for the inclusion of a disadvantaged group—women—and therefore appears to meet our selection criteria. Yet, it was created by Frontier, a public relations and marketing firm that plans and manages clubs and events for the oil and gas industry. Although it presents itself as a women’s advocacy group, this Club is above all a marketing exercise for fossil fuel companies that appear to have adopted women’s advancement to enhance their public image.

We also found many organizations that were easy to classify by goal—they clearly fit our SMO selection criteria—but difficult to classify by domain, most often because their aims mentioned women, women’s empowerment, and/or gender equality even though there was little evidence they actively worked in these areas. Because the UIA reports organizations’ own portrayals of their missions, it assigned these groups to the “Women” category in its subject indices. Listed in this category in the 2013 *Yearbook* is the International Association for the Protection of the Environment in Africa, which aims to

enhance national and regional efforts in protection and improvement of the environment for a better quality of life for the population,

through communication, training, research and counselling . . . promote integration of environmental considerations in the development process . . . encourage participation of women and young people in environment protection.

This organization *does* specifically encourage women’s participation, but this seems a minor concern tacked on to its central mission. Similarly, the UIA also classified as “Women” the Forum Urodynamicum, whose aims are to “advance interdisciplinary research and post-graduate education in the fields of prevention, diagnosis, and treatment of lower urinary tract dysfunction, as well as in female urology.” The word “female” in the organization’s self-description prompted this group’s inclusion in the “Women” subject category, even though the organization neither identifies with nor advocates for women. These are both worrisome examples of how the use of source-defined categories or machine coding of text can generate problematic classifications of organizational domains.

Although some organizations appeared to be paying “lip service” to various domains, other organizations demonstrated a more comprehensive, multi-issue focus. The Women’s International League for Peace and Freedom lists the following as its goals in the 2013 *Yearbook*:

Bring together women who are opposed to war, violence and exploitation and all forms of discrimination and oppression, to unite in establishing peace based on economic and social justice for all; help bring about a world economic and social order founded on absence of violence, . . . on respect for fundamental human rights and on recognition that peace and security depend on equitable sharing of riches; work for rights of women, equality for all, . . . general and complete disarmament, . . . safeguarding the natural environment, respect for international law and the strengthening of the United Nations and its specialized agencies.

We were faced with multiple decisions: is the domain for this organization women’s rights, peace, human rights, development, environment, international integration, or all of these?

Ways Forward

In today's world, classifying organizations correctly requires researchers' careful attention. Yet, prior methodological scholarship has attended mostly to issues of erroneous organizational exclusion. Our examples reveal that mistaken organizational inclusion (or overcounting) is also a problem for researchers seeking to enumerate organizations. Here, we suggest some ways forward.

When coding organizational goals, social movement researchers should consider the ways new organizations may reflect elite appropriation of challenger discourses. Searching out supplemental data sources may be necessary to gauge the true nature of an organization. When coding domains, researchers should not accept publishers' classifications at face value or use keywords to generate counts without additional investigation. Once researchers have a working list of possible organizations of interest, they should review organizational records to determine if the group belongs in their count. Classification challenges will also be reduced when researchers have substantive expertise in the domain of organizations they are attempting to classify. Complementing counts from organizational directories with qualitative analyses of the issues and conflicts reflected in the organizational population will also allow researchers to have a better handle on important boundaries between types of groups and the extent to which discourse has diffused or been appropriated.

For multi-issue groups, there are several approaches to managing organizational complexity. One approach would be to assign such groups to a blanket "multi-issue" category. Although this approach may help researchers home in on remaining organizations working most actively in an area, we do not recommend this approach; it would result in the loss of valuable data about the organization's aims. A second approach would be to create variables for primary and secondary issue foci and assign them accordingly, requiring the researcher to make tough decisions about the relative import of the organization's different concerns. A third approach would be to code each of the organization's foci, preserving more of the organization's complexity at the cost of productive

discriminations. To categorize TSMOs by domain, we used a version of the second approach. We developed a list of about 100 very specific issue categories and another list of about 40 social movement industries. We assigned organizations between one and four issue categories and one social movement industry, allowing us to account for both their specific and general domains.

Of computational methods, supervised learning, which relies on an initial batch of human-coded documents, may be more appropriate for classifying organizations than simpler keyword-based methods. Provided that researchers invest time into the initial batch of documents and carefully supervise the algorithm, the flexibility of supervised learning methods gives them great promise. Keep in mind, however, that both dictionary and supervised methods rely on *a priori* categories, which means that researchers may miss out on new frames or discursive trends.

Conclusions

In sum, the process of assigning organizations to goal and domain categories—indeed, the process of classifying social phenomena in general—has been both facilitated and transformed by source digitization and Web-enabled full-text search, which allows publishers to place organizations in categories algorithmically and offers much touted and inarguable benefits to researchers in terms of efficiency and access. While the growing availability of indexed and searchable organizational databases encourages more large-N research on organizations, we fear that it leads more researchers to overlook the ways these sources can compromise data validity. This is producing more "radically decontextualized research" that confounds scholarly efforts to enhance our understandings of the social world (Putnam 2016:396). Research shortcuts can lead to less valid data. This is true even when human coders are used.

It is impossible to know the extent to which overcounting has already influenced organizational research. Depending on their methods and data sources, researchers may overcount organizational populations to different extents, and the nature of the overcounting itself will

have varied. To be forthright, we suspect that in many cases, overcounting may be unlikely to result in biased correlation and effect estimates. In counts of hundreds or thousands of organizations, overcounting by 5 percent or 10 percent may matter little. But, especially in research that seeks to carve up populations of organizations into smaller categories—distinguishing by their goals or domains—including groups that look like but are not in fact like these organizations could affect our substantive conclusions.

Classifying organizations is an unavoidable part of collecting and analyzing data about them. In developing or using classification systems, researchers should take caution not to abstract their objects of study from the dynamic political, social, and economic environments that shape their forms and frames. In practice, this means being cautious, particularly when relying on keyword search or predefined indices, both of which fail to account for the dynamism of organizations' discursive strategies. It also means being mindful of changes in discourses and organizational fields over time, devoting attention to unclear cases, and complementing the principal data source with additional sources where necessary. The world is complex, and conflict-oriented organizations provide real challenges for those wanting to do systematic research on them. And yet, understanding these particular groups, how they change, and how they relate to other groups over time is essential to explaining how social change happens.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by National Science Foundation (SES-1323130).

ORCID iDs

Samantha Plummer  <https://orcid.org/0000-0003-0331-0349>

Melanie M. Hughes  <https://orcid.org/0000-0002-8724-5355>

Notes

1. Isomorphism refers to the tendency for organizations in similar institutional environments to develop similar structures and practices through mimetic, normative, and/or technical processes (Dimaggio and Powell 1983).
2. The number of SMOs that espouse multi-issue frames in the *Yearbook* has increased in recent decades (Smith and Wiest 2012)

References

- Andrews, K., and B. Edwards. 2005. "The Organizational Structure of Local Environmentalism." *Mobilization: An International Quarterly* 10(2):213–234.
- Andrews, K., B. Edwards, A. Akram, and A.K. Hunter. 2016. "Sampling Social Movement Organizations." *Mobilization: An International Quarterly* 21(2):231–246.
- Bevan, S., F. Baumgartner, E. Johnson, and J.D. McCarthy. 2013. "Understanding Selection Bias, Time-Lags and Measurement Bias in Secondary Data Sources: Putting the Encyclopedia of Associations Database in Broader Context." *Social Science Research* 42(6):1750–1764.
- Bob, C. 2009. *The International Struggle for New Human Rights*. Philadelphia, PA: University of Pennsylvania Press.
- Bromley, P., and J.W. Meyer. 2017. "'They are All Organizations': The Cultural Roots of Blurring Between the Nonprofit, Business, and Government Sectors." *Administration and Society* 49(7):939–966.
- Brulle, R., L. Turner, J. Carmichael, and J. Jenkins. 2007. "Measuring Social Movement Organization Populations: A Comprehensive Census of U.S. Environmental Movement Organizations." *Mobilization: An International Quarterly* 12(3):255–270.
- Burke, M.C., and M. Bernstein. 2014. "How the Right Usurped the Queer Agenda: Frame Co-optation in Political Discourse." *Sociological Forum* 29(4):830–850.
- Carpenter, R. C. 2014. *"Lost" Causes: Agenda Vetting in Global Issue Networks and the Shaping of Human Security*. Ithaca, NY: Cornell University Press.
- DiMaggio, P.J., and W.W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2):147–160.
- De Bakker, F.G., F. Den Hond, B. King, and K. Weber. 2013. "Social Movements, Civil

- Society and Corporations: Taking Stock and Looking Ahead." *Organization Studies* 34(5-6):573-593.
- Edwards, B., and M. Foley. 2003. "Social Movement Organizations Beyond the Beltway: Understanding the Diversity of One Social Movement Industry." *Mobilization: An International Quarterly* 8(1):87-107.
- Fyall, R., M.K. Moore, and M.K. Gugerty. 2018. "Beyond NTEE Codes: Opportunities to Understand Nonprofit Activity Through Mission Statement Content Coding." *Nonprofit and Voluntary Sector Quarterly* 47(4):677-701.
- Goldman, M. 2005. *Imperial Nature: The World Bank and Struggles for Social Justice in the Age of Globalization*. New Haven, CT: Yale University Press.
- Grønbyerg, K. 1994. "Using NTEE to Classify Non-profit Organisations: An Assessment of Human Service and Regional Applications." *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 5(3):301-28.
- Hasenfeld, Y., and B. Gidron. 2005. "Understanding Multi-Purpose Hybrid Voluntary Organizations: The Contributions of Theories on Civil Society, Social Movements and Non-Profit Organizations." *Journal of Civil Society* 1(2):97-112.
- Hughes, M.M., M.L. Krook, and P. Paxton. 2015. "Transnational Women's Activism and the Global Diffusion of Gender Quotas." *International Studies Quarterly* 59(2):357-372.
- Hughes, M. M., P. Paxton, S. Quinsaat, and N. Reith. 2018. "Does the Global North Still Dominate the International Women's Movement? A Network Analysis of Women's International Nongovernmental Organizations, 1978-2008." *Mobilization: An International Quarterly* 23(1):1-21.
- Kim, M., and E. Boyle. 2012. "Neoliberalism, Transnational Education Norms, and Education Spending in the Developing World, 1983-2004." *Law and Social Inquiry* 37(2):367-394.
- Koopmans, R., and P. Statham. 1999. "Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches." *Mobilization: An International Quarterly* 4(2):203-221.
- Martin, A.W., F.R. Baumgartner, and J.D. McCarthy. 2006. "Measuring Association Populations Using the Encyclopedia of Associations: Evidence from the Field of Labor Unions." *Social Science Research* 35(3):771-778.
- Meyer, J.W. 2010. "World Society, Institutional Theories, and the Actor." *Annual Review of Sociology* 36(1):1-20.
- Mooney, P. 2012. "Civil Society Strategies and the Stockholm Syndrome." *Development Dialogue* 61(September):327-331.
- Murdie, A., and D.R. Davis. 2012. "Looking in the Mirror: Comparing INGO Networks Across Issue Areas." *Review of International Organizations* 7(2):177-201.
- Pearce, J. 2013. "Power and the Twenty-First Century Activist: From the Neighbourhood to the Square." *Development and Change* 44(3):639-663.
- Plummer, S., J. Smith, and M. M. Hughes. 2018. "Transnational Advocacy Networks and Global Health Governance." *Global Health Governance* 12(1):62-74.
- Putnam, L. 2016. "The Transnational and the Text-Searchable: Digitized Sources and the Shadows they Cast." *The American Historical Review* 121(2):377-402.
- Reith, N.E., P. Paxton, and M.M. Hughes. 2016. "Building Cross-National, Longitudinal Data Sets: Issues and Strategies for Implementation." *International Journal of Sociology* 46(1):21-41.
- Smith, J., S. Plummer, and M.M. Hughes. 2017. "Transnational Social Movements and Changing Organizational Fields in the Late Twentieth and Early Twenty-First Centuries." *Global Networks* 17(1):3-22.
- Smith, J. and D. Wiest. 2012. *Social Movements in the World-system: The Politics of Crisis and Transformation*. New York: Russell Sage Foundation.
- Wong, W. H. 2012. *Internal Affairs: How the Structure of NGOs Transforms Human Rights*. Ithaca, NY: Cornell University Press.