The Transnational NGO Study: Rationale, Sampling and Research Process

Transnational NGO Initiative, Moynihan Institute of Global Affairs, Maxwell School of Citizenship and Public Affairs, Syracuse University[1]

DRAFT: Last updated on January 18, 2010

Abstract: Transnational NGOs (TNGOs) are increasingly visible and influential actors in global affairs. Although academics and the general public are today more aware of TNGO activity today than a decade ago, our understanding of how TNGO leaders understand their roles in global politics remains limited. Scholarship on this topic has often been driven by individual case studies focusing on successful campaigns within specific sectors. Although research on transnational NGOs is thriving, scholarship largely takes place within particular disciplinary traditions and rarely capitalizes on the strength of an interdisciplinary approach. In addressing these limitations, the TNGO Project focuses on a sample of 152 TNGOs registered in the United States and selected on the basis of size, sector and fiscal health. Researchers traveled the country to question TNGO leaders about governance, accountability, effectiveness, collaboration, communication, leadership and related issues and challenges. The interview protocol was informed by multiple disciplinary perspectives including social movements, not-for-profit management, epistemic communities, transnational activism, interest groups, organizational theory and communications. To facilitate both exploratory (inductive) and confirmatory (deductive) modes of inquiry, the open-ended interview responses were coded using computer aided qualitative data analysis software.

[1] This research was supported by National Science Foundation Grant No. SES-0527679 (Agents of Change: Transnational NGOs as Agents of Change: Toward Understanding Their Governance, Leadership, and Effectiveness) and the Moynihan Institute of Global Affairs at Syracuse University. The TNGO research team included (in alphabetical order) Alejandro S. Amezcua, Tosca Bruno-van Vijfeijken, Derrick L. Cogburn, Jesse D. Lecy, George E. Mitchell, Margaret Hermann, Christiane Pagé, Paloma Raggo, Michael J. Scialdone, Hans Peter Schmitz and Lorena Viñuela.

Introduction

Transnational non-governmental organizations (TNGOs) are increasingly visible actors in global affairs, prompting rapidly expanding academic research across many disciplines in the humanities and social sciences. While non-state actors have always played an important role (Halliday 2001), their numbers and visibility have markedly increased in recent decades. The Union of International Associations (UIA) claims that the number of international NGOs increased from less than 200 in 1909 to more than 20,000 in 2005 (Union of International Associations (UIA) 2005), with much of the growth occurring since the 1970s. Salamon et al. concluded in a 2003 study of 35 countries that civil society organizations represented a $1.7 trillion industry (about 5 per cent of the combined economies). In the United States, Kerlin and Thanasombat identify for the year 2003 a total of 5,600 international not-for-profits with a combined budget of about $17.7bn (Kerlin and Thanasombat 2006).[2]

Transnational NGOs have not only grown in number and capacity, but also with regard to their political influence. The number of NGOs with consultative status at the United Nations increased from a few dozen in 1945 to 3,289 in September 2009 (United Nations Department of Economic and Social Affairs 2009). The World Bank reports that projects with some degree of 'civil society' involvement increased from six percent in the late 1980s to over 70 percent in 2006 (Werker and Ahmed 2008). An expanding literature describes how transnational NGOs shape global governance and influence domestic social and political change (Price 1998; Glasius 2002; Batliwala and Brown 2006; Joachim 2007; Keck and Sikkink 1998; Risse et al. 1999; Koslinski and Reis 2009). The increasing relevance of transnational activism has created much debate

[2] Despite their growth over the past decades, transnational NGOs account for only 2 per cent of the not-for profit sector in the United States.

about the legitimacy and proper role of civil society actors projecting their influence across national borders. While an initial wave of scholarship primarily focused on the potential of a 'global civil society' to advance human rights, environmental protection, conflict resolution and sustainable development,[3] more critical views of transnational activism are today well represented in the literature. Such views challenge the legitimacy and accountability of TNGOs (Anderson 2000; Collingwood and Logister 2005), question the alleged democratizing influence of 'global civil society' (Jaeger 2007), point toward harmful effects and limits of transnational activism (Fisher 1997; Mendelson and Glenn 2002; Kennedy 2004; Schmitz 2006; Carpenter 2007) and identify TNGOs with new forms of Western hegemony (Bob 2005; Chandhoke 2005). Beyond these important debates about the appropriate role of activist organizations, the vast majority of scholars agree that TNGOs, for better or worse, constitute a growing force in contemporary global politics.

Although academics and the general public are today more aware of TNGO activity, our understanding of how TNGO leaders understand their roles in global affairs remains limited in three important ways. First, no large scale systematic studies exist that examine how TNGO leaders define the challenges they confront in carrying out their missions, interacting with states and international institutions, governing their organizations, raising funds to support their activities, responding to accountability pressures, evaluating impact and other issues. Individual or small-N case studies of TNGOs abound, often compiled into anthologies focused on deriving general lessons from a limited number of observations (e.g., Risse and Ropp 1999; Lewis and Wallace 2000; Lindenberg and Bryant 2001; Seckinelgin 2002; Glasius 2002). Indeed, most

[3] Arguably, the establishment of the Global Civil Society Yearbook at the Center for the Study of Global Governance (London School of Economics) in 2001 represents a cumulative endpoint of the then dominant optimistic view regarding the role of transnational activism.

studies of TNGOs focus on a specific sector (e.g., human rights), a single case, or a single issue campaign (e.g., banning landmines). As such, many of the literature's generalizations are based on narrow subsets of TNGOs relying disproportionately on the experiences of organizations considered successful or prominent.

Second, most NGO research is confined to a particular discipline or sub-discipline, such as international relations, public administration, or sociology. Advances in our collective understanding of organized transnational activism have been limited by the persistence of disciplinary divides and segmented research programs. This fragmentation is expressed in the separate debates about TNGOs that have emerged in the literatures on social movements, not-for-profit management, lobbying and interest group behavior, advocacy networks, and epistemic communities. As illustrations of such discipline-driven research, consider publications engaging primarily with Keck and Sikkink's *Activists Beyond Borders* in the field of international relations, the *transnational social movements* literature in sociology (Smith 1997 O'Brien, 2000 #1921), the *epistemic community* literature (Haas 1992), or the efforts to extend the domestic not-for-profit literature to transnational organizations (Salamon and Anheier 1998). By privileging a particular theoretical perspective, each of these literatures only captures a small section of the complex realities confronting TNGOs and their leaders. Since relatively few scholars reach across conventional boundaries to inform their own research,[4] many studies about TNGOs fail to harness the explanatory power offered by interdisciplinary perspectives.

[4] Exceptions of such cross-disciplinary efforts include (Salamon and Anheier 1998; Sikkink and Smith 2001; Tarrow 2005; Ebrahim and Weisband 2007).

Third, rarely do studies of transnational NGOs link an analysis of leadership perspectives to organizational characteristics and outcome measures. Although such analyses are more common in the domestic not-for-profit literature, organizations with a specific transnational focus remain on the margins of academic research. Due to the dearth of appropriate data for large-N studies, we lack a basic descriptive understanding of TNGOs in their role as agents of global change and remain largely ignorant about potentially significant variations across different types of TNGOs. Little is understood about the links between leadership characteristics and organizational attributes, for example, such as whether leaders' definitions of effectiveness or accountability relate to organizational qualities such as size, governance structure, or fiscal health. Moreover, only by exploring how TNGO leaders themselves define and understand such complex constructs as effectiveness or accountability can researchers begin to gage the possible gaps between the ideals and realities of transnational activism.

The transnational NGO study was designed to overcome these three shortcomings. Leaders from 152 TNGOs registered in the United States and spanning five major sectors of transnational activism were interviewed using a protocol informed by six disciplinary perspectives that included questions designed for both exploratory (inductive) and confirmatory (deductive) modes of inquiry. Transnational NGO leaders were asked questions about governance, accountability, effectiveness, collaboration, communication and leadership. Moreover, they led organizations that differed in size, fiscal health and sectoral focus.

A transnational NGO registered in the United States is defined as an organization that (1) has activities in several countries (2) pursues clearly articulated goals, (3) has not-for-profit

501(c)(3) status in the United States, and (4) solicits and receives funding voluntarily from private individuals (and other donors).

SAMPLING

The 152 leaders interviewed for the study were selected from a population of 334 US-based international not-for-profits rated by Charity Navigator (www.charitynavigator.org) in 2005. The Charity Navigator database was chosen because its website was the only source providing information on size, sector, financial efficiency and financial capacity for each organization. These ratings are based on financial information derived from US Internal Revenue Service (IRS) Forms 990. The two major components of the ratings are *financial efficiency*, which measures overhead and fundraising expenditures in relation to program spending, and *organizational capacity*, which assesses the organization's financial stability over the past five years (e.g., measured in cash reserves).[5] The ratings are adjusted for sector and other considerations.

Charity Navigator employs selection criteria that exclude many organizations. Transnational NGOs in the population had 501(c)(3) tax-exempt status from the IRS, had at least four consecutive years of IRS Forms 990 available and had received public support greater than $500,000 during their most recent fiscal years. Organizations that reported zero fundraising costs or that were overwhelmingly funded through government grants or fees for services were excluded from the population, along with private foundations, hospitals, hospital foundations, private universities, colleges, community foundations and public broadcasting stations.

[5] Many not-for-profits have adapted in creative ways to the enhanced focus on overhead spending imposed by external watchdogs. However, since we compiled the sample shortly after the creation of Charity Navigator, we do not expect widespread manipulation of such financial information in order to achieve higher ratings.

A proportionate stratified random sample of 182 cases was drawn from a population of 334 cases rated by Charity Navigator and classified as international.[6] The response rate was approximately 68 per cent as we were able to complete 123 interviews with leaders from the original sample.[7] The final dataset contained 152 organizations, including 29 replacements, which were randomly drawn from the remaining population. In the final sample, 81 percent of respondents were the CEOs, presidents or executive directors of their organizations, 12 percent were vice presidents, 4 percent were lower level directors and 3 percent held other positions.

The strata were defined by five sectoral, three budget (size) and four financial health classifications derived from information provided by Charity Navigator. The five sectors were human rights, sustainable development, conflict resolution, environment and humanitarian relief. The three categories of budget size were small (less than $1 million), medium ($1 million to $10 million) and large (greater than $10 million). The four financial health classifications were based on Charity Navigator's organizational efficiency[8] and capacity[9] ratings. Organizations receiving

[6] The original population contained 35 duplicates that were cross-listed in the database by sector. The sampling procedure drew a proportionate stratified random sample of 182 (177 organizations plus 5 duplicates) from a population of 334 (299 organizations plus 35 duplicates). However, the basis for duplication in the original database was effectively random.

[7] The response rate is calculated as the number of successful interviews from the original sample divided by the number of cases in the original sample or 123/182=0.69.

[8] *Organizational efficiency*, as defined by Charity Navigator: "Analyzing a charity's efficiency reveals how well it functions day to day. Charities that are efficient spend less money to raise more. Their fundraising efforts stay in line with the scope of the programs and services they provide. They keep administrative costs within reasonable limits. They devote the majority of their spending to the programs and services they exist to provide" (from: How do we rate charities?, www.charitynavigator.com, accessed: April 2009).

[9] *Organizational capacity*, as defined by Charity Navigator: "We analyze a charity's capacity to determine how well it has sustained its programs and services over time, and whether it can continue to do so, even if it loses support or faces broad economic downturns. By doing so, we show givers how well that charity is positioned to pursue long-term, systemic change. Charities that show consistent growth and maintain financial stability are more likely to last for years to come. They have the financial flexibility to plan strategically and pursue long-term objectives, rather than facing flurries of fundraising to meet payrolls and other short-term financial obligations" (from: How do we rate charities?, www.charitynavigator.com, accessed: April 2009).

one or two stars on a dimension were classified as low on that dimension and organizations receiving three or four stars were classified as high on that dimension. Table 1 confirms that the sample closely matches the population.

Table 1. Sampling with Replacements

	Population		Sample	
	%	N	%	n
Environment	15%	49	14%	22
Human Rights	14%	48	14%	21
Humanitarian Relief	21%	69	21%	32
Sustainable Development	42%	141	42%	64
Conflict Resolution	8%	27	9%	13
Total		334		152
Small	40%	132	37%	56
Medium	40%	135	42%	64
Large	20%	67	21%	32
Total		334		152
Low Efficiency/Low Capacity	12%	40	14%	22
Low Efficiency/High Capacity	11%	38	9%	13
High Efficiency/Low Capacity	27%	90	23%	35
High Efficiency/High Capacity	50%	166	54%	82
Total		334		152

PROTOCOL DEVELOPMENT

The interview protocol resulted from the interaction of a team of researchers who were interested in applying their various theoretical perspectives to the study of TNGOs and their leaders. The interdisciplinary team brought contemporary questions about the role and significance of TNGOs

from different academic literatures, including international relations (transnational advocacy), social movement studies (sociology), public administration (not-for-profits), social network analysis (collaboration), political psychology (leadership traits) and organizational theory. The protocol was structured around three broad themes: leaders' views regarding (1) their organizations, (2) their leadership, and (3) the transnational aspects of their activities.

Regarding the organizational theme, the team was interested in how TNGO leaders defined their short- and long-term goals, what they saw as examples of success, how they perceived of organizational challenges, how decisions were made, to whom they believed to be accountable, and what their funding strategies were. The team also sought to focus on leadership issues by asking how respondents characterized *good* leaders, how they got into this line of work, how they recruit new personnel, and what the skill sets they perceived as critical to leading their organizations. Finally, the team was interested in understanding the global and transnational aspects of TNGO activities including the impact of transnationalism and their collaborative behavior in networks and partnerships with local communities, governments, corporations, and intergovernmental as well as other non-governmental organizations.

Protocol development began in the fall of 2004 and was completed in the spring of 2006, shortly before researchers began interviewing. During the period of development, researchers were engaged in extensive consultations with practitioners to refine the broad areas of inquiry for the study and later to pilot test draft versions of the protocol.

In May 2005, small number of TNGO leaders and outside researchers were invited for a retreat to central New York in order to solicit their input. In February 2006, a larger group of TNGO leaders was consulted at a workshop held in Washington, DC. A draft protocol was developed and pilot tested on leaders of Southeast Asian and Indian NGOs and further revised to reflect their feedback. Following these various inputs, the protocol was deliberately designed with open-ended questions in order to fully capture how TNGO leaders speak in their own words about the core challenges of organizational effectiveness, accountability, and leadership. We realized in the pretests that offering pre-defined answers and forcing choices would have limited the capture of important perceptions and insights by those leaders.

The interviewing process began in July 2006 and ended in April 2008. During this time, several minor changes were made in response to the early incoming interview responses, including more detailed inquiries into the role of boards and funding (questions added in October 2006). Other modest revisions were made to the protocol in October 2007 after the research team had begun reviewing completed transcripts. A set of questions about communication tools was moved to the front of the protocol in order to improve the flow of the interviews. Two questions on transnationalism, which had initially been placed between the effectiveness and accountability sections, were also moved to the front. A handout was created for a series of questions about organizational activities to improve respondent recall and a set of repetitive questions on collaborations and partnerships was consolidated. The final interview protocol is available at [http://www1.maxwell.syr.edu/moynihan/tngo/Welcome/]

INTERVIEWING PROCESS

Interviews were administered face-to-face at leaders' preferred locations, usually their offices, to ensure the interview contexts would be comfortable and appropriate to the subject matter. The nature of respondents' high-level positions also required us to be as accommodating and flexible as possible with regard to their demanding time schedules.

The official written requests for interviews asked for 90 minute time slots. Most respondents initially committed to between 30 and 60 minutes. However, the actual interviews lasted much longer as respondents became comfortable and reflective about issues they rarely had an opportunity to discuss in detail. Interviews averaged 82.5 minutes; the shortest was 32 minutes and the longest was 153 minutes. Interviewers ultimately collected a total of approximately 209 hours of digital recording (12,537 minutes). The recorded interviews were professionally transcribed and subsequently coded.

The interview process occurred over two phases. In the first phase, 12 graduate students conducted 28 interviews during the summer of 2006. Prior to interviewing, students familiarized themselves with relevant scholarly literature on NGOs and participated in meetings with faculty, including the PIs, to discuss the materials. Students were also asked to complete an online human subjects training course offered through the Syracuse University Institutional Review Board (www.citiprogram.org). Most students already had practical experience with NGOs or with elite interviewing. Students underwent intensive training to improve their comfort with the interview protocol. They also practiced with their peers over a period of three weeks, during which marked improvements were observed in memorization and ease. Each student was then subjected to

simulation exercises with faculty and experienced staff. Aiming to achieve high quality data collection, students were paired-up in the field, with one student conducting the interview and the other recording the digital audio and taking notes.

In phase two, the process was streamlined to enable a smaller number of interviewers to build greater expertise more rapidly. From October 2006 to August 2008, the remaining 124 interviews were conducted by two Ph.D. students, the project administrator (an experienced former TNGO practitioner) and one faculty member. When possible, interviewers were matched with respondents sharing similar interests to facilitate rapport.

Face-to-face interviews present specific challenges, including (1) reactivity as a result of distorting effects due to the interview situation, (2) interviewer bias or the clouding of responses with personal opinions or preconceptions of the interviewer, and (3) respondent bias in the form of giving answers driven by perceived expectations, faulty recall, or a desire to withhold relevant information (Padgett 2008). To enhance respondent candor, we assured participants full confidentiality and worked deliberately on strategies to foster participant-interviewer rapport. Interviewers also completed debriefing forms in which they assessed respondent candor and other issues that may have affected the quality of the interviews. These debriefing forms were appended to the interview transcripts for coding. Where this debriefing information was available, data show that 86 percent of respondents were perceived as very candid, 14 percent evinced occasional lack of candor and none displayed prolonged lack of candor. Instances of occasional lack of candor involved hesitation at discussing issues that were currently confidential within the organization such as succession planning, delayed recall, the impact of a leader's age

on the pace of the interview, and telephone and staff interruptions. In no instance was occasional lack of candor judged to warrant discarding data.

CODING AND DATA STRUCTURE

The principal investigators selected Atlas.ti for manual coding because of its capabilities for organizing and quantifying large amounts of qualitative information. The software also had features facilitating the collaboration of larger research groups. This section describes how Atlas.ti was used to create both qualitative and quantitative datasets.[10]

The transcripts were preprocessed before coding. A project document (PD) was created for each interview. These PDs were prefaced by demographic and technical information and appended by debriefing notes. The PDs were loaded into hermeneutic units (HUs) within Atlas.ti and coded separately by a team of five coders. Each coder's HU contained a complete list of codes organized into 'code families.' Generally, code families correspond to specific interview questions and the codes within them represent responses.

During coding, coders saw a list of code families in the left pane of the program's graphical user interface (GUI) and the text of a PD in the main window to the right. When a leader responded to an interview question, coders exploded the corresponding code family, reviewed the available codes and selected the appropriate text or 'unit of meaning' for coding. Coders created 'quotations' by dragging and dropping codes onto these selections of text. Quantitatively, the

[10] For an online introduction to using Atlas.ti (and other qualitative software), see various primers posted at *Online QDA. Learning Qualitative data Analysis on the Web* (at: http://onlineqda.hud.ac.uk/index.php). See also the CAQDAS Networking Project at the University of Surrey (http://caqdas.soc.surrey.ac.uk/).

creation of a quotation generates a count of one for the corresponding cell in the 'frequency count report' exported from Atlas.ti.

To facilitate coding, the arrangement of code families in the left hand pane of the GUI followed the sequence of the protocol. The master codebook was further organized into nine major sections, 38 subsections, 91 code families and 413 individual codes following the general structure of the protocol. The code labels reflected this structure. For example, the code '1021_pat_poh_ceo' refers to the codebook section numbered 1000, subsection for personal attributes, code family for position in organizational hierarchy and code for CEO. The variable labels were generated from the unique numerical prefixes.

Each code belongs to one of five different types of code families. The five types, which are summarized in Table 2 below, are descriptive, frequency, latent, index and exploratory. The codes within these families were designed (and applied) to capture qualitative information in a manner conducive to quantification. This design produced both a rich qualitative dataset containing retrievable quotations and a quantitative dataset containing a numerical summary of the qualitative information.

Descriptive code families contain groups of codes used primarily to capture attributes of respondents and their organizations. Codes in these families were designed to be collapsed into variables with mutually exclusive categories. Coders chose only one code from each family to apply to a particular quotation, but once chosen could apply that code multiple times within a PD. For example, a coder may have applied the code '1021_pat_poh_ceo' to an instance of a

leader affirming his or her job title. If the respondent further elaborated upon his or her professional responsibilities later in the interview, this may also have been coded with the same code. In the quantitative dataset, these data would have been transformed into a value of '1' indicating 'CEO' for the variable 'v1020,' 'Position in Organizational Hierarchy.'

Frequency code families are the default type and characterize most of the codes in the codebook. These code families are similar to multiple-response variables which are not mutually exclusive. Coders could have chosen multiple codes from a frequency code family and applied them to any number of quotations within a PD. For example, a coder may have applied the codes '2021_ogs_gst_edu' for 'Education' and '2022_ogs_gst_adv' for 'Advocacy' from the code family 'Organizational Goals and Strategies' to several overlapping quotations within a PD. In the quantitative dataset, the variables 'v2021' for 'Advocacy' and 'v2022' for 'Education' would show the counts for each of these codes for the corresponding PD. For convenience, these counts have been dichotomized in the binarized dataset.

Latent code families contain codes that capture latent content interpreted by coders. Codes in these families were designed to be collapsed into variables with mutually exclusive categories. Coders could have chosen one code from a family and then applied it multiple times within a PD. For example, a coder may have developed the impression that the pressures facing the respondent's organization for greater accountability were mainly internal. The coder would have applied the code '5141_pal_pre_int' from the code family 'Source of Accountability Pressures Acting upon the Organization' to the leader's responses to various accountability questions. In the quantitative dataset, this would have been transformed into a value of '1' for the categorical

variable ‘v5160’ of the same label. Latent code families are logically identical to descriptive code families, except that the former capture latent content whereas the latter capture manifest content.

Index code families contain codes that are mutually exclusive at the level of the quotation but that are not mutually exclusive at the level of the project document. Codes in these families were intended to be used to create ratio or index variables. There is only one index code family, 7100, for ‘Collaboration Structure.’ For example, a coder could have applied the code ‘7121_col_str_age’ for ‘Respondent’s organization is the agent in the relationship’ to five quotations and the code ‘7122_col_str_pri’ for ‘Respondent’s organization is the principal in the relationship’ to two different quotations. In the quantitative dataset, researchers could use the two corresponding count variables, v7121 and v7122, to create a new variable indicating that the respondent’s organization was ‘usually the agent’ in collaborations. The index code family is similar to frequency families except that coders were prohibited from applying more than one code from the family per quotation.

Exploratory code families contain codes that highlight qualitative information not directly intended for quantitative analysis. Codes in these families were intended to be used in an explorative manner for qualitative research using Atlas.ti. While nearly all codes were designed to be used in this way, codes from exploratory code families are essentially uncategorized and may not be meaningful for quantitative analysis. In the quantitative dataset, data for these variables simply indicate how many times coders applied the respective codes.

The table below summarizes the five types of code families, the types of variables into which they have been converted and the instructions given to coders for each type of family. The column for 'codes' indicates whether the family appears as a single response or multiple response variable in the quantitative dataset. The 'quotation' column indicates whether coders were instructed to apply codes only once or as many times as needed to a single quotation. The 'document' column indicates whether coders were instructed to apply codes only once within a PD or as many times as needed.

Table 2. Code Families as Variables

Code Families as Variables					
Code Family	**Variables**	**Codes**	**Quotation**	**Document**	**Instructions**
'Descriptive'	Nominal Ordinal	Mutually Exclusive	Mutually Exclusive	Mutually Inclusive or Exclusive	Select only one; code only once or code each instance
'Frequency'	Count	Mutually Inclusive	Mutually Inclusive	Mutually Inclusive	Select all that apply
'Latent'	Nominal Ordinal	Mutually Exclusive	Mutually Exclusive	Mutually Inclusive	Select only one; code each instance
'Index'	Ratio (index)	Mutually Inclusive	Mutually Exclusive	Mutually Inclusive	Code each instance accordingly
'Exploratory'	Count	Mutually Inclusive	Mutually Inclusive	Mutually Inclusive	Select all that apply

Coders were required to successfully pass the frequency count report from each coded transcript through an error-checking tool before submitting their completed 'copy bundles' for inclusion in the master hermeneutic unit. The error-checking tool checked for errors of omission and logical contradictions among nominal and ordinal values. For each code family, it returned either a message highlighted in green indicating no errors, a message in yellow indicating a potential error of omission or a message in red indicating either a logical contradiction or an error of

omission where a value was required. This information was consistently and effectively used to identify and correct mistakes made during the coding process.

DATA TRANSFORMATION

All individual hermeneutic units were merged into a Master Hermeneutic Unit from which two types of data were exported. The first type of data consists of quotations listed by organization and code. The second type of data was exported as a frequency count report. The frequency count report was merged with secondary data and imported into Stata 10 (http://www.stata.com/) for transformation. A Stata do file was run to transform code families with mutually exclusive values into categorical or ordinal variables as appropriate. The new variables were labeled based on the original code family names. Multiple response variables from frequency code families and exploratory code families were only labeled. Each value of each variable also received a label. Stata Transfer was used to convert the Stata data files into multiple formats.

The frequency count report exported from Atlas.ti is simply a table of counts. If a code was not applied within a particular transcript, the cell was assigned a value of zero. Some zero values indicate that constructs were not salient to respondents while other zero values indicate missing data. The Stata do file recoded zeros as missing when no responses were coded for an entire section of a respondent's interview. Additionally, missing data can be identified for individual interview questions by converting zeros to 'missing' if no codes were applied within the code families.

INTERCODER ALIGNMENT

Intercoder alignment is the process by which coders come to achieve agreement on the presence or absence of specific constructs in specific interview transcripts. Intercoder agreement is a state in which coders agree with one another in their coding choices given the same evidence.

To ensure that the coding scheme was as complete and explicit as possible, the original technical codebook provided many descriptors and examples to clarify the meaning of various codes. Contentious codes were defined through a continuous process in which virtually every code was exhaustively discussed by the entire coding group and its meaning fixed by consensus. At least one PI was present during each major qualitative alignment meeting and in nearly all cases explicitly approved any significant definitional modifications.

Coding exercises were designed with the involvement of the coders and the PIs to collectively align on definitions, units of meaning and coding procedures. The exercises also addressed the specific challenges of recognizing and making contingent coding choices. This led to important enhancements to the structure of the codebook during the initial phase of coder orientation.

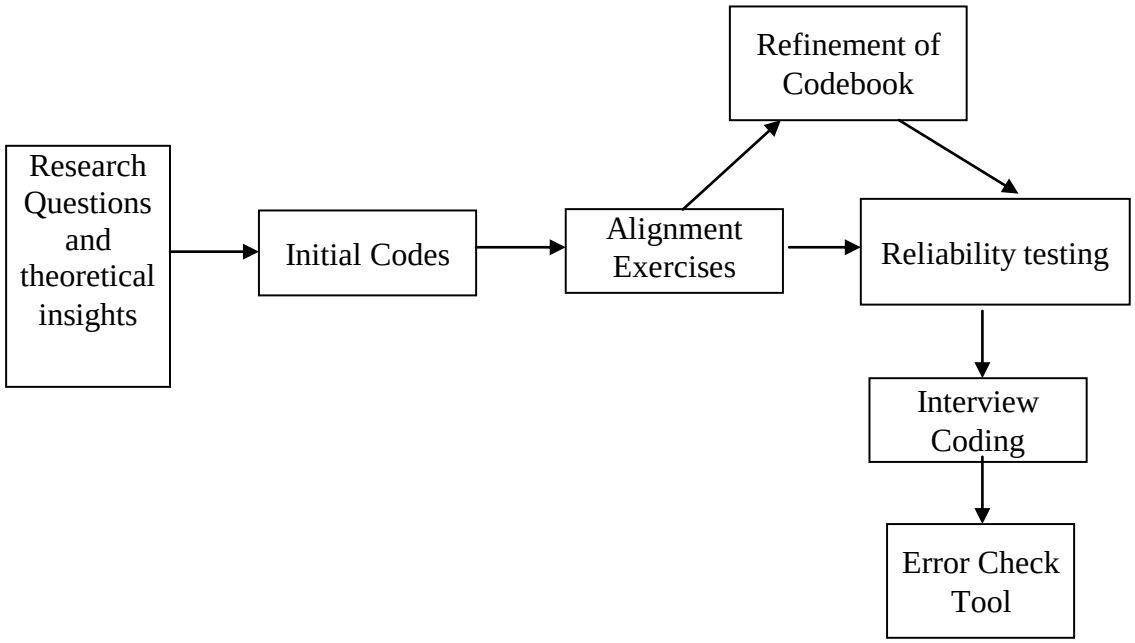
During the coder orientation process, coders met with the PI and co-PIs to review, restructure, and refine the initial coding scheme into a hierarchy comprised of sections, sub-sections, code families and specific codes. The first round of major revisions encompassed mostly structural and logical changes. Subsequent revisions focused mainly on qualitative alignment. Intimate coder and PI participation in the revisions of the codebook greatly improved intercoder

alignment and even helped promote group camaraderie, which we found to be invaluable for a project of this magnitude and complexity.

Our diagnostic activities systematically assessed and improved intercoder agreement using qualitative and quantitative techniques. Conventional scores of intercoder reliability were inappropriate because they compare codes applied to identical units of text. With Atlas.ti, however, units of text (quotations) are delimited at the discretion of individual coders. Intercoder reliability scores must be calculated at the level of the project document rather than at the level of the quotation because quotations differ across coders.

The process of intercoder alignment involved detailed and extensive exercises to improve the application of the codebook, enhance intercoder agreement, identify and correct logical errors and produce meaningful measures of intercoder reliability. The overall alignment process is summarized in Figure 1 below, which illustrates the various steps of this process. Such processes are particularly useful for the systematic development of codes applicable to open-ended interviews (Mayring 2000).

Figure 1. Open-ended interview coding and alignment



The initial coding exercises during the summer of 2006 were designed to familiarize coders with the codebook, to discover miscellaneous problems or errors and to identify points of intercoder disagreement. For the first alignment exercise, coders each coded a single practice interview independently and then met over a period of several days to compare their results code-by-code to identify problems and disagreements and to reach greater concurrence. We exported frequency count data for each code by coder and created a simple spreadsheet formula to score each code with a measure of convergence/divergence. We reviewed each code and compared the quotations of text each coder highlighted to support his or her inference. We used this process to align on the presence or absence of codes and also on quotation delimitation. We observed, qualitatively, a high degree of agreement on nominal code families and an adequate degree of agreement on ordinal code families. In most cases, this exercise revealed that coders tended to agree with high consistency on the delimitations of units of meaning.

Each diagnostic activity led to conceptual clarifications and to the enumeration of specific coding criteria to enhance intercoder reliability. We also analyzed frequency count data, word count data, and various other quantitative scores, such as word count per code, for each coder to aid group alignment on quotation delimitation and code application frequencies.

After the completion of several coding exercises designed for orientation, we conducted three full scale intercoder reliability tests on three different out-of-sample interviews. The first test transcript was coded by each coder independently, the second test transcript was coded by a dyad and a triad of divergent coders and the third test transcript was coded by each coder independently. After each test, data were exported from the qualitative software and the results were systematically and exhaustively compared and discussed, usually over periods of several days.

In the latter diagnostic activities, although a small number of new definitions and examples were added, our attention turned to a few persistent areas of intercoder disagreement, which mostly stemmed from difficulties in coding ‘latent’ code families and from omitted code bias. Most ‘latent’ code families captured latent content reflecting coders’ subjective interpretations of textual evidence, while ‘frequency count’ code families simply captured manifest content. To enhance intercoder reliability for these peculiar code families, it was especially important for coders to achieve intersubjective agreement on codes within ‘latent” families. The alignment exercises helped to resolve evident coder disagreement through construct clarification, consensus

building and criteria enumeration. A PI with a particular expertise in the substantive area of inquiry was usually also consulted to resolve any remaining issues.

To improve intercoder alignment on 'latent' families, we undertook frequent exercises to promote group convergence on the coding of latent content. The challenge arises naturally as coders scanned multiple levels of content and observed contradictions inherent in the rhetoric, orientation and perceptions of respondents. Given, for instance, that the typical TNGO president or CEO has a long-term vision in mind that he or she sometimes articulates in terms of short-term steps, responses can reflect partially factual information and partially the aspirations of the respondent. Such textual evidence may contradict itself, strictly speaking. In our group discussions of these difficulties, we considered whether we had 'one-off' disagreement (1 vs. 4) or balanced disagreement (e.g., 2 vs. 3) among coders. In the first case, the minority presented his or her reasons for his or her code and the rest of the group responded by adducing their own reasons for their other code. Final determinations were made by consensus and usually favored the majority. In the second case, each coder adduced his or her justifications for their code and the group unanimously agreed upon which reasons were legitimate—adopting those collectively agreed upon and discarding the rest. This often led to the enumeration of specific coding criteria, many of which were added as descriptors in the codebook. Further discussions additionally led to agreement upon new best practices, such as how long to mull over a quotation before coding it as 'insufficient content.' To discover the evidential roots of coder disagreement, we queried Atlas.ti on the relevant code families and reviewed each coder's quotations and debated individuals' justifications until consensus was achieved.

We also converted our test frequency data into binary data to produce strictly presence/absence datasets. In doing this, we effectively treated each transcript as a single quotation, despite the fact that each transcript normally contains more than a hundred quotations. This data reduction enabled us to test quantitatively for intercoder reliability at the level of presence/absence for each code for each interview transcript, while ignoring the qualitative content actually highlighted by the coders. We constructed overall concurrence scores for each coder by comparing each coder to each of his or her peers and then to the scores of the coder group collectively. The formula we employed counted the number of coders with whom the coder was in agreement contingent upon whether the coder coded presence or absence. We then averaged the two scores separately for each coder.

Using the quantitative diagnostics described above, we were able to identify individual divergences from the group, which we reduced by paring dyads and triads of coders with relatively low agreement rates. These dyads and triads then coded another practice interview, facilitating alignment between divergent coders, and then they collectively compared their results, facilitating group alignment. After each iteration, we discussed disagreements by code and family, clarified definitions and coding criteria and made further revisions to the codebook, if necessary. We converged on quotation lengths, coding frequencies and coding rules, *inter alia*.

In addition to numerous alignment and diagnostic activities, coders were assigned to sit together, side-by-side, during the first eight weeks of coding. Constant peer questions, discussions and active debates quickly came to be the norm. Real-time dialogue and consensus-building enabled substantive issues to be decided quickly and agreed upon simultaneously by all coders.

Particularly contentious issues were flagged by the coding team leader and were presented to a PI for final adjudication.

For the purposes of alignment activities, we identified five types of intercoder disagreement including nominal (categorical), ordinal (degree), omission (applying too few codes), commission (applying too many codes), and other (disagreements involving 'other' codes); of these, nominal and ordinal disagreements are the first and second most significant, respectively. We found that our most significant substantive disagreements arose from the coding of latent content. We believe this difficulty to have been caused by 1) the ambiguous and often contradictory nature of many responses and 2) the idiosyncratic nature of the inferences required for interpreting and categorizing textual evidence. Table 3 below summarizes the different types of intercoder disagreements.

Table 3. Explanations for Intercoder Disagreements

Type of disagreement	**Explanation of disagreement**
Nominal	Coders disagree over the appropriate qualitative code for a particular quotation; could be a matter of degree, depending on definitions of codes; important for qualitative and quantitative analysis
Ordinal	Coders disagree over the value of the ordinal variable; important for latent content; disagreement is a matter of degree; important for quantitative analysis
Omission	Undercoding; coders fail to code when sufficient textual evidence is present; important for frequency counts and qualitative analysis
Commission	Overcoding; coders apply codes for which sufficient textual evidence is lacking; inappropriate delimitation of units of meaning; introduces error
Other	Disagreement concerning applications of 'other' codes is relatively benign, provided that they are used appropriately; underuse of 'other' codes causes informational inefficiency qualitatively; disagreement could be indicative of over- or under-confident coding

While the procedures discussed above were effective for promoting intercoder alignment, they were not ideal for producing descriptive measures of intercoder agreement. For this, ten in-sample interviews were coded twice by separate coders. One coded transcript from each pair was arbitrarily chosen to be included in the final dataset. The 413 codes were converted into 263 variables with which to calculate intercoder agreement. Some codes were appropriately combined into their nominal, ordinal or interval variables, while codes used for qualitative analysis, such as 'other,' were ignored. Reflecting our experience with our complex coding scheme, intercoder agreement was calculated separately for three logically distinct types of variables. Intercoder agreement scores were calculated for each of the ten pairs of coders and then averaged. Values logically range from zero, indicating complete intercoder disagreement, to one, indicating complete intercoder agreement. Among 25 'demographic' nominal and ordinal variables, intercoder agreement was 0.74. Among 225 'frequency' count variables agreement was 0.82. Among 13 'latent' nominal and ordinal variables agreement was 0.55. The overall weighted intercoder agreement score was 0.80.

CONCLUSIONS

The TNGO study is limited in several significant respects. First, the results can only be generalized to TNGOs with entities registered in the United States as 501(c)(3) tax-exempt organizations. Second, every organization had to be rated by Charity Navigator, which may have resulted in other biases limiting the representativeness of our sample. Third, statistics derived from our data are subject to both sampling and coding error. We have gone at great lengths in minimizing both sources of error. The relatively large sample size (152 out of 334) provides also some assurance against the undesirable effects of sampling and coding error.

Notwithstanding these limitations, the TNGO study is among the first to collect original, detailed primary data from a representative sample of TNGO leaders and their organizations. The TNGO data reveal how leaders conceptualize important issues that have either been garnered sufficient attention by scholars or have been studied without an explicit recognition of leadership perspectives on those crucial topics. The methodological pluralism and combined leadership and organizational perspectives embodied in the TNGO study enable longstanding empirical questions about governance, effectiveness, or accountability to be addressed. Moreover, the study offers new opportunities for the formulation of novel research questions heretofore both unasked and unanswered.

The implementation of the research design can be understood with reference to three core principles. The first component of the triad is *practicability*. The project had to be completed within budget and the coding scheme had to be comprehensible and manageable. Coders spent about half a day coding each interview. More extended coding periods would not have increased coding validity. The second component is *epistemological balance*. The codebook and coding procedures were designed to facilitate both deductive, confirmatory inquiry and inductive, exploratory inquiry. In the process of coding and interpreting the data, many tradeoffs between these two orientations emerged. A purely deductive approach would have required a completely static codebook with no opportunity for emergent coding and the discovery of new conceptualizations. A purely inductive approach would have precluded statistical generalizations since not all interviews would have been coded according to the same schema. Additionally, hypotheses could not have been formally tested. Finally, in developing a continuously charting a course between deductive and inductive reasoning, we established the third principle of

informational efficiency. Optimizing the research design for one or the other type of inquiry would not have taken full advantage of the large sample size together with the informational richness of the transcripts. Thus, the code families and their attendant coding procedures were designed to extract the maximum amount of combined qualitative and quantitative information.

The result of this pragmatic triad of practicability, epistemological balance and informational efficiency is a pair of complementary datasets that enables researchers to swing seamlessly between inductive and deductive modes of inquiry to form information-rich, yet widely generalizable observations about transnational NGOs and their leaders.

Scholars have paid increasing attention to transnational NGOs, their goals, strategies and increasingly apparent impact. Accumulating knowledge about the role of TNGOs in global affairs has been hampered by disciplinary boundaries, the prevalence of small-N studies and a lack of research designs capable of revealing links between leadership, organizational characteristics and organizational behavior. This study sought to overcome these limitations by implementing a design consisting of a large number of in-depth interviews with leaders from a representative sample of TNGOs using an interview protocol developed by an interdisciplinary team of researchers specializing in Political Science, International Relations, Sociology, Public Administration, Political Psychology and Information Studies. The combination of primary data collected directly from TNGO leaders and secondary data collected about their organizations provides new opportunities for understanding TNGO leadership, governance and effectiveness.

WORKS CITED

Anderson, Kenneth. 2000. "The Ottawa Convention Banning Landmines, the Role of International Non-Governmental Organizations, and the Idea of International Civil Society." *European Journal of International Law* 11 (1): 91-120.

Batliwala, Srilatha, and L. David Brown, eds. 2006. *Transnational Civil Society. An Introduction*. Bloomfield: Kumarian Press.

Bob, Clifford. 2005. *The Marketing of Rebellion: Insurgents, Media, and International Activism*. Cambridge: Cambridge University Press.

Carpenter, R. Charli. 2007. "Setting the Advocacy Agenda: Theorizing Issue Emergence and Nonemergence in Transnational Advocacy Networks." *International Studies Quarterly* 51 (1): 99-120.

Chandhoke, Neera. 2005. "How Global is Global Civil Society?" *Journal of World-Systems Research* 11 (2): 354-71.

Collingwood, Vivien, and Louis Logister. 2005. "State of the Art. Addressing the INGO 'Legitimacy Deficit'." *Political Studies Review* 3: 175-92.

Ebrahim, Alnoor, and Edward Weisband, eds. 2007. *Global Accountabilities. Participation, Pluralism, and Public Ethics*. Cambridge: Cambridge University Press.

Fisher, William F. 1997. "Doing Good? The Politics and Anti-Politics of NGO Practices " *Annual Review of Anthropology* (26): 439-64.

Glasius, Marlies. 2002. "Expertise in the Cause of Justice. Global Civil Society Influence on the Statute for an International Criminal Court." In *Global Civil Society 2002*, ed. M. Glasius, M. Kaldor and H. Anheier. Oxford: Oxford University Press, 137-68.

Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination." *International Organization* 46 (1): 1-35.

Halliday, Fred. 2001. "The Romance of Non-State Actors." In *Non-State Actors in World Politics*, ed. D. Josselin and W. Wallace. Houndmills: Palgrave, 21-40.

Jaeger, Hans-Martin. 2007. ""Global Civil Society" and the Political Depoliticization of Global Governance." *International Political Sociology* 1 (3): 257-77.

Joachim, Jutta. 2007. *Agenda Setting, the UN, and NGOs. Gender Violence and Reproductive Rights*. Washington, D.C.: Georgetown University Press.

Keck, Margaret E., and Kathryn Sikkink. 1998. *Activists Beyond Borders. Advocacy Networks in International Politics*. Ithaca: Cornell University Press.

Kennedy, David. 2004. *The Dark Sides of Virtue. Reassessing International Humanitarianism*. Princeton: Princeton University Press.

Kerlin, Janelle A., and Supaporn Thanasombat. 2006. "The International Charitable Nonprofit Subsector. Scope, Size, and Revenue." Washington D.C.: Urban Institute.

Koslinski, Mariane Campelo, and Elisa P. Reis. 2009. "Transnational and Domestic Relations of NGOs in Brazil." *World Development* 37 (3): 714-25.

Lewis, David, and Tina Wallace, eds. 2000. *New Roles and Relevance: Development NGOs and the Challenge of Change*. Hartford, CT: Kumarian Press.

Lindenberg, Marc, and Coralie Bryant. 2001. *Going Global. Transforming Relief and Development NGOs*. Bloomfield, CT: Kumarian Press.

Mayring, Philipp. 2000. "Qualitative Content Analysis." *Forum: Qualitative Social Research* 1 (2).

Mendelson, Sarah E., and John K. Glenn. 2002. *The Power and Limits of NGOs*. New York: Columbia University Press.
Padgett, Deborah K. 2008. *Qualitative Methods in Social Work Research*. Los Angeles, CA: Sage.
Price, Richard. 1998. "Reversing the Gun Sights: Transnational Civil Society Targets Land Mines." *International Organization* 52 (3): 613-44.
Risse, Thomas, and Stephen C. Ropp. 1999. "International Human Rights Norms and Domestic Change: Conclusions." In *The Power of Human Rights. International Norms and Domestic Change*, ed. T. Risse, S. Ropp and K. Sikkink. Cambridge: Cambridge University Press, 234-78.
Risse, Thomas, Stephen C. Ropp, and Kathryn Sikkink, eds. 1999. *The Power of Human Rights. International Norms and Domestic Change*. Cambridge: Cambridge University Press.
Salamon, Lester M., and Helmut K. Anheier. 1998. "Social Origins of Civil Society: Explaining the Nonprofit Sector Cross-Nationally." *Voluntas. International Journal of Voluntary and Nonprofit Organizations* (9): 213-48.
Schmitz, Hans Peter. 2006. *Transnational Mobilization and Domestic Regime Change. Africa in Comparative Perspective*. Houndmills, Basingstoke: Palgrave Macmillan.
Seckinelgin, Hakan. 2002. "Civil Society as a Metaphor for Western Liberalism." *Global Society* 16 (4): 357-76.
Sikkink, Kathryn, and Jackie Smith. 2001. "Infrastructures for Change. Transnational Organizations, 1953-1993." In *Restructuring World Politics. Transnational Social Movements, Networks, and Norms*, ed. K. Sikkink, J. V. Riker and S. Khagram. Minneapolis: University of Minnesota Press, 24-44.
Smith, Jackie. 1997. "Characteristics of the Modern Transnational Social Movement Sector." In *Transnational Social Movements and Global Politics*, ed. J. Smith, C. Chatfield and R. Pagnucco. Syracuse: Syracuse University Press, 19-58.
Tarrow, Sidney. 2005. *The New Transnational Activism*. Cambridge: Cambridge University Press.
Union of International Associations (UIA). *International Organizations and NGOs*. Union of International Associations (UIA), 2005 [cited. Available from http://www.uia.org/statistics/pub.php.
United Nations Department of Economic and Social Affairs. *Consultative Status with ECOSOC and other accreditations*. United Nations 2009 [cited. Available from http://esango.un.org/civilsociety/.
Werker, Eric, and Faisal Z. Ahmed. 2008. "What Do Nongovernmental Organizations Do?" *Journal of Economic Perspectives* 22 (2): 73-92.