# Coding Sub-State Actors using the CAMEO (Conflict and Mediation Event Observations) Actor Coding Framework *

Philip A. Schrodt, Ömür Yilmaz, Deborah J. Gerner and Dennis Hermrick

Dept. of Political Science

University of Kansas

Version 1.0B1 : March 20, 2008

## Abstract

Because almost all contemporary conflicts transcend the traditional focus on state actors, featuring instead significant involvement of both sub-state and non-state actors, the state-centered coding schemes used in older data sets such as WEIS and COPDAB have proven inadequate for coding current events. In their place, we have established a systematic method of hierarchically creating codes that allow for the identification of states, sub-state actors, ethnic groups, geographical regions, IGOs and NGOs. This system has proven sufficient to code a wide range of relevant actors involved in inter- and intra-state protracted conflicts in Africa, the Balkans, Central Asia and the Middle East.

# 1 Introduction

One of the major changes in the post-Cold War environment has been the emergence of sub-state actors as major forces in both domestic and international politics. Many have argued that the proliferation of sub-state, non-state, multi-state, and trans-state actors has blurred almost completely the traditional separation of "international" and "comparative" politics, a theme that generally motivates our efforts. At times these groups exercise coercive force equal to or greater than that of states, whether from within, as in the case of "failed states", or across borders, as with Israel's attempts to control Hizbollah in Lebanon and Hamas in Gaza, or the near irrelevance of borders in many of the conflicts in central and western Africa. Irrespective of the effectiveness of their coercive power, these non-state actors may also be a source of identity that is more important than that of an individual's state-affiliation—the ability of al-Qaeda to attract adherents from across the Islamic world is a good example—or provide examples of strategies that are imitated across borders, as has been seen in the numerous non-violent popular revolutions in Eastern Europe.

In order to study such actors, we need to be able to coherently code them in a manner that allows groups to be compared over time and geographical region. For several decades, two coding frameworks dominated event data research: Charles McClelland's WEIS (1976) and the Conflict and Peace Data Bank (COPDAB) developed by Edward Azar (1982). Both were created during the Cold War and assumed a "Westphalian-Clausewitzian" political world in which sovereign states reacted to each other primarily through official diplomacy and military threats. While innovative when first created, these coding systems are not optimal for dealing with contemporary issues such as ethnic conflict, low-intensity violence, organized criminal activity, and multilateral intervention. McClelland (1983: 177) viewed WEIS as only a "first phase"; he certainly did not anticipate that it would continue to be used, with only minor modifications, for four decades.

Because they were state-centered, WEIS and COPDAB paid relatively little attention to non-state actors. A small number of long-lived opposition groups that were active in the 1960s such as the Irish Republican Army, the Palestine Liberation Organization, and the National Liberation Front of Vietnam (Viet Cong) were given state-like codes, as were major international organizations such as the United Nations and the International Committee of the Red Cross/Red Crescent. From the perspective of coding, these actors were treated as honorary states. Beyond this small number of special cases, sub- and non-state actors were ignored.

A major breakthrough in the systematic coding of sub-state actors came with the Protocol for the Analysis of Nonviolent Direct Action (PANDA) project in the early 1990s, which introduced the concept of sub-state "agents"—e.g. "media", "politicians", "labor unions"—as part of their standard actor coding. PANDA's primary focus was on contentious politics within states, and consequently needed to distinguish, for example, between police and demonstrators, or between government and opposition political parties. The PANDA actor–agent system has been extended into the Bond et al.'s (1997) Integrated Data for Event Analysis (IDEA) system (`http://vranet.com/idea/`; also see King and Lowe, 2003).

Unlike PANDA, which coded the entire world, the Kansas Event Data System (KEDS) project (`http://web.ku.edu/keds`) has focused specifically on regions that have experienced protracted conflicts. As a consequence, rather than using the PANDA/IDEA of introducing new agent fields, we initially maintained the WEIS/COPDAB convention of using a single "source" and "target" field. However, because the areas we are coding involved quite a few sub-state actors, we began to develop a series of standard codes that were a composite of the WEIS nation-state codes concatenated with PANDA agent codes. Under this system, for example, ISRMIL would be "Israel military", "LIBOPP" would be "Liberian opposition parties, "SIEGOV" would be Sierra Leone government and so forth.

In 2004, we received a contract from the United States government's multi-agency Political Instability Task Force (`http://globalpolicy.gmu.edu/pitf/`) to develop five-year data sets on 18 areas that were at risk for mass killings. In developing dictionaries to code these areas, several problems in our existing protocols became apparent in addition to the state-centered legacy of the WEIS codes.

First, we found that in the fifteen-year development of automated coding dictionaries, we hadn't been completely consistent in the creation of sub-state codes, particularly when distinguishing ethnic groups located in multiple states, as frequently occurred in the Balkans. Second, there were conceptual inconsistencies even at the level of our primary codes that had been based on WEIS. For example, the United Nations had its own three-character code, UNO, but other inter-governmental organizations were coded with codes ranging from three to six characters depending on their common abbreviation—for example IMF, OSCE, UNHCR, UNESCO, ECOWAS—or simply the generic ING for "international group."

More generally, we needed a set of rules that an individual doing dictionary development could apply when she encountered a new political actor to generate an appropriate code that could be easily interpreted later. These rules have subsequently been incorporated into the CAMEO—Conflict and Mediation Event Observations—event coding scheme.

# 2   General Principles

Three principles underlie the CAMEO actor coding system. First, codes are composed of one or more three-character elements: In the present system a code can consist of one, two or three of these elements (and therefore three, six, or nine character codes), although this may be extended later. These code elements are classified into a number of broad categories, such as state actors, sub-state actor roles, regions, and ethnic groups.

Second, the codes are interpreted hierarchically: The allowable code in the second element depends on the content of the first element, and the third element depends on the second. This is in contrast to a rectangular coding system, where the second and third elements would always have the same content. The most familiar analogy to a hierarchical coding system is the Library of Congress cataloguing system, where the elements of the catalog number vary—systematically—depending on the nature of the item being catalogued, and consequently may

contain very different information despite being part of a single system. The event coding system used in the Behavioral Correlates of War data set (Leng 1987) is another example of a hierarchical scheme in the event data literature. This hierarchical system is the key aspect that differentiates CAMEO's composite codes from the PANDA/IDEA actor-agent system, which is rectangular.

Third, we are basing our work on standardized codes whenever these are available. This is most obvious in our use of the United Nations nation-state codes (ISO-3166-1 ALPHA 3). `http://unstats.un.org/unsd/methods/m49/m49alpha.htm` gives a list of these codes and `http://www.iso.org/iso/en/prods-services/iso3166ma/04background-on-iso-3166/index.html` gives details on the process. This contrasts to the Russett-Singer-Small (1968) codes used in WEIS, which are specific to the North American international relations community. We have generally adopted the IDEA agent codes for sub-state actors, we've used the HURIDOCS (`http://www.huridocs.org/`) classifications for world religions, and we may incorporate ISO-3166-2 codes for urban and regional divisions (for example, cities, states, and provinces) when these are available.

Unfortunately, standard codes are generally *not* available. For example, most IGOs are known by acronyms of varying lengths, so we need to decide how to truncate these to three characters. We spent considerable time trying to determine whether the U.S. government had a standard list of militarized non-state actors; as best we can tell, this does not exist (or at least not in a form we can access).

The Minorities at Risk project (MAR; `http://www.cidcm.umd.edu/inscr/mar/`) would seem to be a logical source for codes for ethnic groups and does have three-character codes for almost 300 groups, but by design MAR is only interested in ethnic groups that are at risk during the 1945-2005 period, rather than all ethnic groups. In particular, a country's dominant ethnic group—which will generally be neither a minority nor at risk—will not appear in MAR unless it is at risk in another country. Thus, for instance, while the CIA World Factbook (CWF)(`http://www.odci.gov/cia/publications/factbook/index.html`) lists seven ethnic groups in Senegal (Wolof 43.3%, Pular 23.8%, Serer 14.7%, Jola 3.7%, Mandinka 3%, Soninke 1.1%, European and Lebanese 1%), MAR tracks only one. In Afghanistan the CWF lists seven groups (Pashtun 42%, Tajik 27%, Hazara 9%, Uzbek 9%, Aimak 4%, Turkmen 3%, Baloch 2%); MAR lists four. MAR and CWF do not always disagree at this level—for example, their lists for Russia and Nigeria are similar—but in general MAR does not appear to be a comprehensive source. That said, we have no particular investment in our set of codes. The dictionaries can always be changed in a few hours with a simple search-and-replace operation and as additional standards become available we could easily adopt them.

# 3   Common Rules, Format, and General Codes

## 3.1   First Three-Characters

### 3.1.1   Country Codes

In the case of domestic actors—actors associated with specific countries—the first three characters indicate the respective countries. The United Nations list of standard three-letter country codes are used to identify countries. The current list, as well as a list of changed and added codes, can be found at the UN website `http://unstats.un.org/unsd/methods/m49/m49.htm`. This list, in turn, forms the basis for a standardized list of country identification codes maintained by the International Organization for Standardization [sic] (ISO) ISO-3166-1 `http://www.iso.org/iso/en/ISOOnline.frontpage`. [1]

While ISO-3166 codes are an international standard, they are not without some problems. While many of the codes are familiar English-language mnemonics—USA is the United States, EGY is Egypt, and ISR is Israel—others are transliterations of non-English state names: DZA is Algeria, CHE is Switzerland, and BIH is Bosnia. This is only a minor problem during dictionary development, since the coders quickly learn the appropriate codes, but the data are somewhat difficult to skim.

Second, the ISO-3166 codes are subject to politics, conspicuously in the absence of a code for Taiwan. Those codes are also a product of the 1970s and consequently codes for entities that existed earlier, such as French Northwest Africa, have to be created if events are generated for historical periods. [2] The ISO-3166-3 standard tracks changes in the coding scheme since the origin of the standard in 1974, but an explicit decision was made not to try to deal with name changes prior to that time; the UN list at `http://unstats.un.org/unsd/methods/m49/m49chang.htm` only goes back to 1990 and consequently, unlike the Russett-Singer-Small (1968) codes, does not include North and South Vietnam or East and West Germany as distinct codes. Similarly, in cases of civil war, a decision needs to be made when to change the coding of a geographical area from a sub-state region to a country, a choice with significant political implications. What does one

---

[1]ISO-3166-1 also provides a list of two-character country codes; these are the codes that are used as "top-level-domain" abbreviations in Web addresses. The ISO does not provide the list of ALPHA-3 codes on the Web, but the UN code list is on the web, and the codes are readily available at other sites such as `http://www.nationsonline.org/oneworld/countrycodes.htm` and `http://www.unc.edu/ rowlett/units/codes/country.htm`.

If space were at a premium, the two-character codes would be sufficient to provide unique identifiers. However, they have even less mnemonic value that the ALPHA-3 codes—for example ICELAND = IS; INDIA= IN; INDONESIA=ID; IRAN = IR; IRAQ = IQ; IRELAND = IE; ISRAEL = IL. For that same reason, there seems little point in using the UN Statistical Office numerical codes.

[2]We are currently not doing this, and at the present time the absence of suitable machine-readable sources makes it unlikely that anyone would do so using automated methods. However, to the extent that one might want to compare automated sets to data that were human-coded from historical sources it could become an issue.

do with the Western Sahara? Tibet? East Timor prior to its independence? The unification of Yemen? Whatever is done, the decision is not "value neutral."

### 3.1.2  Religious/Ethnic Identity Codes

Typically individuals acting in their own capacities or on behalf of their countries are given the codes that correspond to their respective countries. However, some ethnic and religious identity groups are not associated exclusively with particular countries. These groups, and hence individuals associated primarily with these groups, are assigned their own special codes. For example, Albanians live in and are significant actors not only in the state of Albania but in other Balkan countries as well; therefore, when news reports specifically mention ethnic Albanians and not the state of Albania, we distinguish between the two by assigning the code ABN as opposed to ALB, which corresponds to Albania.

In addition to the special identity codes, there are also codes for more universal religious groups, such as CHR for Christian, CHRMRN for Maronite Christian, MOS for Muslim, and MOSSUN for Sunni Muslim. See Table 1 for a list of these religious group codes. These religious and ethnic identity codes are used as the first three-characters *only* when the actor in question is not specifically identified with a country in the news lead. Otherwise, they become the second three-characters (i.e., suffixes to country codes). For example, "Kurds" are coded as KUR, while "Iraqi Kurds" are coded as IRQKUR.

### 3.1.3  Generic International/Transnational Actor Codes

Different generic codes are used to differentiate between various kinds of international and transnational actors. IGO (international governmental organization), NGO (non-governmental organization), NGM (non-governmental movement), and MNC (multi-national corporation) are these generic codes. These generic codes can either be used on their own or as the first three characters of more detailed codes. Some international/transnational actors get their own special three character codes (e.g. UNO for the United Nations, AMN for Amnesty International, IRC for the Red Cross), which are not used on their own but as suffixes to these generic actor codes (i.e. IGOUNO, NGOAMN, NGOIRC).

As an exception, we also have a six-character generic code used for peacekeeping forces when the particular organizational affiliation is not known: IGOPKO. This code is assigned even when the national identity of the peacekeepers in question is specified. Hence, for instance, "Senegalese peacekeepers" are coded as IGOPKO since they operate as part of an inter-governmental organization and they might be representing the United Nations or ECOWAS.

In some cases, actors are primarily transnational/international in nature, yet their country affiliations are also known; coders have the option of including this information by attaching country codes to the generic transnational/international codes. This could be particularly

Table 1: Main Religious Group Codes

| Religious Group | Code |
|---|---|
| Agnostic/Atheist | ATH |
| Alewi | MOSALE |
| Animist/Pagan | PAG |
| Bahai | BAH |
| Buddhist | BUD |
| Christian | CHR |
| Catholic | CHRCTH |
| Coptic | CHRCPT |
| Jehovah's Witnesses | CHRJHW |
| Latter Day Saints | CHRLDS |
| Maronite | CHRMRN |
| Orthodox Christian | CHRDOX |
| Protestant | CHRPRO |
| Confucian | CON |
| Hindu | HIN |
| Jain | JAN |
| Jew | JEW |
| Hasidic | JEWHSD |
| Orthodox/Ultra-Orthodox Jew | JEWUDX |
| Muslim | MOS |
| Druze | MOSDRZ |
| Shi'a | MOSSHI |
| Sufi | MOSSFI |
| Sunni | MOSSUN |
| Taoist | TAO |
| Zoroastrian | ZRO |

valuable if, given the research agenda, the country distinction becomes key at the analysis stage. (For example, actors with codes NGOUSA, NGMUSA, and MNCUSA, for instance, can then all be combined with other USA actors at this stage, while still preserving the full codes/information in the dictionaries for alternative groupings.)

In addition, we have the code UIS (unidentified state actor), which is used when an actor is known to be a country or government—or it is known to act on behalf of a country or state—but the identity of the particular country is not revealed in the report (e.g. "foreign diplomat"). UIS is typically used as a three-letter code on its own.

The distinction between NGO and NGM is meant to capture the theoretical difference between well-structured, formal non-governmental organizations and anomic or non-associational social movements. Although in many instances the line dividing the two will be fuzzy, we believe that the distinction is theoretically important enough-it might be more important for some research questions than others-to warrant having separate codes. Greenpeace, for instance, is one of those difficult cases: although it is typically thought to be an NGO, it actually functions more as a loose and informal movement with some more formal organizations, such as the Greenpeace Foundation and Greenpeace USA, associated with it.

Sometimes news articles refer to unnamed actors such as "human rights advocates," "anti-WTO protesters," and "supporters of Palestine;" these actors are best coded as NGMs since they clearly belong to some non-governmental collective effort but, at the same time, are not explicitly associated with specific organizations. "Aid workers," on the other hand, are coded as NGOs since participation in aid distribution generally requires being part of an organized group with the necessary resources to distribute material assistance—even if the identity of the group is not specified in the news lead.

Table 2 illustrates what these generic codes refer to and their usage with examples. Note that some of these examples are simply assigned the three character generic codes, while others are further specified with their own special suffixes. In some cases this depends on how specific the report is with respect to the identity of the actor. However, the level of specificity could also vary from project to project, depending on what kinds of actors are important given the research question, and hence, how much information coders want to retain in the code for each actor.

### 3.1.4   Geographic Region Codes

Sometimes news reports do not specify a group of countries separately and instead refer to them using the general geographical region they are associated with, such as Latin America (LAM), the Middle East (MEA), Eastern Europe (EEU), etc. In such cases, where exact identification of the countries involved is not possible, international region codes laid out in Table 3 can be used as the first three characters, which then typically constitutes the entire code.

Table 2: International/Transnational Generic Codes

| Generic Code | Actor Type | Example | Full Code |
|---|---|---|---|
| IGO | International or regional Inter-governmental organization | "the United Nations" "World Trade Organization" | IGOUNO IGOWTO |
| INT | International or transnational actors who cannot be further specified as IGO, UIS, NGO, NGM, or MNC | "international envoy" "international observer" "world community" | INT INT INT |
| MNC | Multi-national corporations | "Halliburton" "multinational firm" "Shell oil company" | MNC MNC MNC |
| NGM | Non-governmental movements | "Greenpeace" "anti-WTO activists" "human rights advocate" | NGMGRP NGM NGM |
| NGO | Non-governmental organizations | "aid worker" "Amnesty International" "Red Cross" | NGO NGOAMN NGOIRC |
| UIS | Unidentified state actors | "foreign diplomat" "world governments" | UIS UIS |

## 3.2   Second Three-Characters

### 3.2.1   Generic Domestic Actor/Role Codes

Generic domestic codes are assigned to actors in order to indicate their roles and statuses, when known and relevant, within their respective countries (see Table 4). These codes—such as GOV for government, OPP for opposition, and REB for armed and violent (non-state) groups—are used as the second three-characters, appended to either country or group identity codes based on the rules described above. In case of REB, note that association with a particular state (for instance, LBNREB, ISRREB, etc.) does not necessarily indicate violent opposition against that state, but only that the group is located and operates from the given country.

A comprehensive list of these generic role codes can be found in Table 4. We make a crucial distinction between primary, secondary, and tertiary role codes: Primary codes are those that identify the role of a domestic actor if at all possible; among those, GOV, MIL, OPP, and REB are in fact the most commonly used. If none of the primary codes applies to the actor in question, the secondary role codes are used. Hence, for instance, a labor union would have the LAB code and a given journalist would have the MED code *only* if they cannot be identified as OPP.

Table 3: International Region Codes

| Region | Code |
|---|---|
| Africa | AFR |
| Balkans | BLK |
| Caribbean | CRB |
| Caucasus | CAU |
| Central Africa | CFR |
| Central Asia | CAS |
| Central Europe | CEU |
| East Indies | EIN |
| Eastern Africa | EAF |
| Eastern Europe | EEU |
| Europe | EUR |
| Latin America | LAM |
| Middle East | MEA |
| Mediterranean | MDT |
| North Africa | NAF |
| North America | NMR |
| Persian Gulf | PGS |
| Scandinavia | SCN |
| South America | SAM |
| South Asia | SAS |
| Southeast Asia | SEA |
| Southern Africa | SAF |
| West Africa | WAF |
| "the West" | WST |

Although we have a code for the legislative branch (LEG), it is identified as a secondary code and is used sparingly: When a given legislative body is mentioned as an organization (e.g. the parliament, the House of Commons, the Senate), LEG may be used. When a particular political party or individual member of the legislature is in question, however, GOV or OPP is used, depending on whether the relevant party has control of the executive branch.

Tertiary role codes, on the other hand, are used only as last resort. RAD captures ambiguous identifiers such as "radical," "extremist," and "fundamentalist" which can be encountered in news reports but do not refer to any systematically identifiable group or role. We felt compelled to create the code to systematize the the coding of such ambiguous labels, the meaning of which could vary from reporter to reporter and across regions: Does the term "extremist" refer simply to the conservative nature of a group or does it imply that the group in question is armed and violent? In order to avoid bias and to ensure reliability, RAD (and not REB) is used in such cases. For example, "extremist Serbian nationalist" is coded as SERRAD. Similarly, MOD is used when ambiguous identifiers such as "moderate" and "mainstream" are encountered.

UAF are also used as a last-resort when an armed group cannot be identified either as MIL or REB. This would most likely arise in cases where the association of a given armed group with the state it operates in is unclear (i.e. whether it is an independent rebel group or a paramilitary). If the link between a paramilitary and a state is common knowledge, however, MIL would still be used—even though the group might not officially be part of the state military institution. The Serb Volunteer Guard, also known as Arkan's Tigers, for instance, is coded as SRBMIL.

In some cases countries are divided up into distinct and politically relevant regions, typically in federal or otherwise decentralized systems, and different codes are assigned for each of these sub-state regions, which become the second three character codes attached to the country codes. In these cases, role codes for domestic actors become the last three character codes (following country and region codes).

It is conceivable that in some cases it would be useful to preserve more than one generic role code that is applicable to an actor in question, as in the case of "state-media" or "military courts." When this happens, the primary role codes take precedence and are used as the second three-characters; it is optional to use the other role code as the third three-characters. Hence, "state-media" could be coded as XXXGOVMED. In case of "military courts," both MIL and JUD are primary codes but MIL takes precedence as it is more imperative that we catch the 'military vs. civilian' distinction (i.e., XXXMILJUD).

### 3.2.2   Sub-State Region Codes

In countries with federal systems, autonomous regions, other forms of decentralization, or any other idiosyncratic facts that render regional distinctions politically significant, our codes link actors to sub-state regions as well as countries. Assigning actors domestic region

Table 4: Generic Domestic Role Codes

| Primary Role Codes | Description |
|---|---|
| COP | Police forces, officers |
| GOV | Government: the executive, governing parties, coalitions partners |
| JUD | Judiciary: judges, courts |
| MIL | Military: troops, soldiers, all state-military personnel |
| OPP | Political opposition: opposition parties, individuals, activists |
| REB | Rebels: armed and violent (non-state) groups, individuals |
| SPY | State intelligence, secret service |
| Secondary Role Codes | Description |
| BUS | Business: businessmen, companies, etc. |
| CVL | Civilian individual or group |
| EDU | Education: educators, schools, students |
| ELI | Elites: former government officials or celebrities |
| LAB | Labor: workers, unions |
| LEG | Legislature: parliaments, assemblies, "lawmakers" |
| MED | Media: journalists, newspapers, television stations,etc. |
| REF | Refugees |
| Tertiary Role Codes | Description |
| MOD | Moderate: "moderate," "mainstream," etc. |
| RAD | Radical: "radical," "extremist," "fundamentalist," etc. |
| UAF | Armed forces that cannot be identified as MIL, COP, or REB |

codes (as the second three characters) allows researchers to code and study intrastate events which might have domestic as well as international significance. Serbia during 2003-2006, for example, is assigned the code SCGSRB, where SCG is the UN code for the state of Serbia and Montenegro and SRB denotes the Republic of Serbia, which is a sub-state entity within Serbia and Montenegro.

In the regional West African actors dictionary, we also have region codes within Liberia and Nigeria, both of which were coded separately and extensively as parts of independent projects we completed during 2003. The dictionary includes separate six character codes for all 36 states of Nigeria—such as NGAABU for Abuja, NGAKAD for Kaduna, and NGALAG for Lagos—whereby the first three characters indicate the country of Nigeria and the second three characters specify the particular state within the federal structure.

In some cases, we have assigned geographic regions within a country their own three character codes because the distinction was important for demographic or other political reasons (even though these regions did not have legal status). Biafra, which is coded as NGABIA, and the Niger Delta Region, coded NGANDR, are two such examples from Nigeria. In case of Turkey, we have given Southeast Turkey its own code (TURSOE), which has allowed us to capture many domestic events (particularly between Kurdish insurgents and the Turkish state) we would not have been able to do otherwise.

### 3.2.3 Religious/Ethnic Identity Codes

These codes refer to ethnic or religious identity groups which are not strictly associated with single countries, thereby requiring their own three character codes. As previously discussed, these codes are assigned as the first three character codes when the report does not link the identity group in question to a specific location or country (e.g., SER for ethnic Serbs when not further specified). When such groups are specifically identified as residing in particular countries, however, religious/ethnic identity codes are added to the country codes (e.g., HRVSER for ethnic Serbs living in Croatia). When the first three characters identify the country and the second three characters indicate a relevant sub-state region or political entity, the identity codes become the last three characters, such as in the case of SCGKSVSER, which refers to Kosovar Serbs. These identity codes can in some cases also be composed of six letters (or two three-character codes). For example, the code for the Druze is MOSDRZ, and when an actor is specifically known to be a Lebanese Druze, then the code becomes LBNMOSDRZ.

### 3.2.4 Special International/Transnational Actor Codes

International/transnational actors can be assigned special codes by attaching actor specific codes as suffixes to the generic codes, particularly to IGO, NGO, NGM, and MNC (i.e. IGOUNO, NGOAMN, NGOIRC). Table 3.6 in Schrodt and Yilmaz (2007: 107-108) lists about sixty such actors for whom we have assigned special codes in our regional dictionaries;

both regionally and globally relevant actors are listed, but note that this list need not be final and we anticipated that other coders/researchers will give additional actors their own codes. If an NGO, NGM, or MNC is known to be identified with a particular country, this information can also be preserved in the code.

### 3.2.5   Country Codes

In cases where the actor in question is identified as an NGO, NGM, or MNC, this role takes precedence over the question of what country they are associated with. Hence, the first three characters become NGO/NGM/MNC and the country code becomes the second three characters *if* the coder/researcher wishes to preserve that information. Attaching the country code does not indicate that the actor is officially identified with or that he acts on behalf of that state. However, it does allow the analyst to eventually collapse together all actors associated with a certain country if s/he so chooses.

## 3.3   Third Three-Characters

### 3.3.1   Generic Domestic Actor/Role Codes

Generic domestic codes indicate the roles and status of actors within their respective countries. As discussed above, these codes become the last three characters of an actor code when the second three characters refer to either the identity group or the sub-state region. For example, the Albanian Army for National Liberation is assigned the code MKDABN-REB, where the first three characters indicate the country (the Former Yugoslav Republic of Macedonia), the second three characters indicate the ethnic group (Albanian), and REB designates the particular group in question as a rebel group.

### 3.3.2   Religious/Ethnic Identity Codes

As mentioned in the previous section on identity codes, in cases where the first three characters identify the country and the second three characters indicate the sub-state region or political entity, the religious/ethnic identity codes—such as TRK for Turks, CRO for Croats, MOS for Muslims, and HAU for the Hausas—are appended as the last three characters. SCGSRBABN for ethnic Albanians living in Serbia, and NGABIAIBO for Ibos living in the Biafra region of Nigeria are two such examples.

### 3.3.3   Branches of International Organizations

In some cases it may be useful to retain information regarding specific branches of international organizations—such as the International Atomic Energy Agency of the United Nations

(UNO)—in the actor code. In such cases, the branch organization can be given a special three-character code that becomes the last three characters of the actor code (e.g., the International Atomic Energy Agency would be coded IGOUNOIAE).

### 3.3.4   Special Actor Codes

In cases where the researcher can—given the amount of information available in the news lead—and wants to distinguish between different actors of the same generic domestic role, different groups are each given their own three-character codes, which are then be used as the last three-characters. For example, the Likud and Meretz Parties in Israel are assigned the nine character codes of ISRGOVLKD or ISROPPLKD and ISRGOVMRZ or ISROPPMRZ, respectively.

# 4   Regional Dictionaries

At this point in the CAMEO Project we have three main regional dictionaries—the Middle East, the Balkans, and West Africa. We have also developed a unique, separate dictionary for Turkey. In addition to following the same format and rules, these dictionaries have a chunk of entries—actors and corresponding codes—in common. Most countries and major international actors, for instance, are found in all of the dictionaries. They differ from each other only in that each contains additional entries that are relevant only for the issues and the countries in that particular region; the difference occurs because we develop separate dictionaries—verbs and actors—for each region using leads relevant for that region. However, because the creation of dictionaries is systematic and consistent process, the regional dictionaries can be compared and merged at any time to build comprehensive main dictionaries. (We merge our dictionaries periodically; hence, the initial dictionaries we use in our regional dictionary developments were at one time created from the merging of a number of smaller and more regional dictionaries.) The following sub-sections describe some of the special actor codes—those that have special suffixes attached to make them more specific than the generic codes, as well as the special group identity codes—that are found in respective dictionaries.

Even in the regional dictionaries, in many cases domestic actors are simply assigned generic codes (such as country or identity code plus the domestic role code) when they are entered in the dictionaries. It is only when one wants to make a distinction between different actors with the same generic code—for instance, between two or more coalition partners in a government—that special codes are created.

## 4.1   The Middle East

The Arab-Israeli conflict, particularly the conflict between Israel and the Palestinians, has been our primary focus in coding the Middle East. Hence, our dictionary is most developed with respect to domestic political actors in Israel and Palestine. As part of a separate project, we have also extensively coded Algeria, and the actors dictionary from that project has been merged with our Levant dictionary. We started coding Turkey using this more general Middle East dictionary. Due to the presence of idiosyncratic codes, we have not merged the actors dictionary we developed for Turkey back into this dictionary; this could still be done, however, selectively.

Table 3.7 in Schrodt and Yilmaz (2007: 112) shows a list of the actors with special codes in the Middle East dictionary. Note that because of the dynamic nature of the domestic positions of many of these actors (for instance, an opposition party yesterday but a government coalition party today), many are date-restricted, so that the domestic generic codes that specify their positions can vary depending on the date of each news report. 'd.r.' refers to code that make extensive used of the 'date-restricted' option in the TABARI coding program, which allows different codes to be assigned to a phrase depending on the date of the news report.

While the code PSE (UN code) refers to the Occupied Palestinian Territories of the West Bank and the Gaza Strip, PAL refers to Palestinians as an identity group. Therefore, Palestinian government and other state actors are coded as PSEGOV, PSECOP, etc. depending on their respective roles. The Palestinian Liberation Organization (PLO), however, is coded as PALPLO as it represents the Palestinian people in general; because of its unusual status, the PLO could not be assigned one of the generic domestic role codes. The organizations underneath the PLO are each assigned codes based on their spheres of influence. Hence, Yasser Arafat and Fatah are date restricted as PALPLO before the Oslo Accords of 1993-which marked the establishment of the Palestinian Authority-and as PSEGOV thereafter. (Fatah itself is in fact further specified as PSEGOVFTA.)

## 4.2   Turkey

CAMEO's actor dictionary for Turkey is unique in that it includes idiosyncratic codes which require that it not be used without certain modifications when coding other countries. The reason is that it gives vaguely identified actors codes that identify them as actors associated specifically with Turkey (even if the information is not explicit in the new leads). For example, "rebels" who are not further specified are coded as TURREB and "soldiers" similarly unspecified are coded as TURMIL. We were able to do that since Turkey was the only country included in that project and it was the only country listed in our search string for news leads; we could reasonably assume that the insufficiently specified actors we come across would in most cases be associated with Turkey.

Table 3.7 in Schrodt and Yilmaz (2007: 114) shows a list of these unique codes which are

present only in our special actors dictionary for Turkey. Note that because TABARI gives precedence to longer patterns over less specified, shorter entries, when news leads further specify the identity of these actors and they are entered into the dictionary as such, these idiosyncratic codes are trumped. For example, "Kurdish rebels" are coded as KURREB and "Iraqi police" is coded as IRQCOP.

In the context of Turkey, "village guards" refer to Kurdish locals in Southeast Turkey recruited by the military to fight Kurdish guerrillas, mainly the rebels associated with the Kurdistan Workers' Party (PKK)–coded TURREBPKK. Hence, although they are not officially associated with the Turkish state and the Turkish military, village guards are coded as TURMIL; this is how we would code state-sponsored paramilitaries.

We also deviate from the normal CAMEO protocol in coding what are called "State Security Courts" (Devlet Guvenlik Mahkemeleri, DGMs) in Turkey. Until June 1999, these courts included military judges and were commonly regarded as being controlled by the Turkish Armed Forces. Therefore, although court systems are typically coded as JUD (or TURJUD in case of Turkey), we code "State Security Courts" as TURMIL for dates prior to 990618. These courts were finally abolished in May 2004.

Unlike Table 3.7, Table 3.8 in Schrodt and Yilmaz (2007: 114) presents a list of special actor codes which could be integrated into other dictionaries. Again, note that what is listed here are the *codes* that one encounters in the dictionary or in the output at the analysis stage, and not all the entries that correspond to each code. More than with many of our dictionaries—though Israel is similar in this respect—coding for Turkey is dependent on date-restrictions; with the exception of post-2003, the multi-party period in Turkey has been marked by short-lived coalition governments and frequent elections. Furthermore, the banning of political parties and their rebirths with slightly changed names have been commonplace in Turkey, thereby requiring date-restrictions to null-code closed parties.

## 4.3   West Africa

In addition to coding West Africa in general with CAMEO, we also coded Liberia and Nigeria for separate projects. Therefore, the West Africa dictionary is most developed for these two countries. Also, because of the level of political decentralization and the importance of intra-state ethnic interactions in these countries, numerous special region codes have been developed particularly to represent the federal states within Nigeria and the counties of Liberia. The domestic region codes for Nigeria are listed in Table 3.10 in Schrodt and Yilmaz (2007: 116) and those for Liberia are listed in Table 3.11. Note that some of these regions, such as the Niger Delta Region in Nigeria, do not constitute legal boundaries but still represent politically important divisions within their respective countries.

Other region-specific special codes are listed in Table 3.12 in Schrodt and Yilmaz (2007: 117). Ethnic groups are coded as six character codes if the exact location of the actors is not specified: the Ogoni people of Nigeria, for example, are coded as NGAOGO as long as the news report does not associate the actor with a more specific region within the country,

but as NGAABUOGO if the specific actor in question is identified as being from Abuja. The same applies to other ethnic groups and other regions. Note that most of the political and militant groups and organizations in this region are not being assigned special codes; this is not because they are not important enough to warrant special codes, but because the first six characters are typically the codes for the country and the region/the ethnic group (NGAHAU, NGAAGU, LBROGO, etc.), and the last three characters are then generally used to specify the roles of the actors. For example, the O'odua Peoples Congress (a Yoruba rebel group) in Nigeria is assigned the code NGAYRBREB-Nigeria, Yoruba, rebel group-which is not a special code (i.e., any Yoruba rebel group would be assigned the same code). Also, note that only a few of the special political organization codes are date-restricted; this reflects the current state of the dictionary, but this could, and should, change as coding continues and the roles of these actors change.

## 4.4   The Balkans

Our focus in coding the Balkans has primarily been on the conflict and conflict resolution events during the first half of 1990s. The Balkans actors dictionary is, therefore, most developed with respect to ethnic and territorial divisions (as opposed to specific political parties or organizations).

The state of the Socialist Federal Republic of Yugoslavia disintegrated by 1992 with the breaking away of its constituents republics, eventually forming the states of Slovenia (UN code SVN), Croatia (HRV), the Former Yugoslav Republic of Macedonia (MKD), Bosnia and Herzegovina (BIH), and Serbia and Montenegro (SCG).

Bosnia and Herzegovina, created with the Dayton Agreement of 14 December 1995 which brought three years of civil war to an end, has a federal structure that consists of two republics: the Bosniak/Croat Federation of Bosnia and Herzegovina (BIHBHF) and the Bosnian Serb Republica Srpska (BIHSRP). Generic role codes (such as GOV, MIL, OPP, etc.) for actors associated with BIHBHF and BIHSRP become the last three characters of the actor codes. In order to differentiate between the states/republics and the people as ethnic groups, Bosnian Muslims are coded as BIHMOS (not BIHBHF), Bosnian Croats as BIHCRO, and Bosnian Serbs as BIHSER (not BIHSRP). More generally, CRO and SER refer to Croat and Serb ethnic groups. If an actor with a given ethnicity is associated with either one of the federal units specifically, the ethnicity code can be attached to the six-character unit code (e.g. BIHBHFSER). Generic role codes (such as GOV, MIL, OPP, REB) can also be used as the last three characters.

After the fellow constituent republics of Slovenia, Croatia, Macedonia, and Bosnia-Herzegovina declared independence in 1991, Serbia and Montenegro-the remaining federal states of the Socialist Federal Republic of Yugoslavia-formed the Federal Republic of Yugoslavia (then UN code FRY). On February 4, 2003, however, a new constitution was accepted, abdicating this self-proclaimed successor to the Socialist Federal Republic of Yugoslavia and replacing it with a loose federation called Serbia and Montenegro (UN code,

hence the CAMEO code, becomes SCG). The new federation consisted of the two states of Serbia (SCGSRB) and Montenegro (SCGMTN), as well as the two autonomous provinces of Kosovo (SCGKSV) and Vojvodina (SCGVVD). With Montenegro's unilateral declaration of independence on 3 June 2006, followed by Serbia's declaration on 5 June 2006, SCG also ceased to exist and gave way to two independent states–MTN and SRB (with SRBKSV and SRBVVD as autonomous provinces).

Note that the state of Serbia has a code that is different from that of the ethnic group of Serbs, who might or might not be living in Serbia. When an actor is associated with the ethnic group of Serbs and its country of origin is not specified, the actor is assigned the code SER; if the Serb in question is associated with a certain location such as Bosnia-Bosnian Serb-then the code becomes BIHSER (not BIHSRB or BIHSCG). The same rule applies to the other ethnic groups.

A more comprehensive list of major actor groups in the region and their respective codes can be found in Table 3.13 in Schrodt and Yilmaz (2007: 120). Note that actors with generic codes are not listed; the point here is to document the different codes—not to list all actor entries, which would mean replicating the whole dictionary-t—hat exist in the dictionaries so that codes which show up in the output can be easily identified during analysis. Actors listed in the last group are derivations of different ethnic groups living in different countries. This list is not exhaustive; those listed here are meant as examples of how state/country codes and codes for ethnic groups living in those countries are merged to create special group codes.

# 5   Future Work

Prior to the advent of automated coding, the development of event data involved a great deal of speculation about the likely nature of the actor and event codes that would best describe political activity. Because human coders typically produce between five and ten events per hour, and a large data set contains tens of thousands of events, experimental recoding was simply not feasible, leading to years or even decades of "lock-in" once those decisions had been made.

Automated coding, in contrast, allows researchers to experiment easily with alternative coding rules that reflect a particular theoretical perspective or interest in a specific set of issues. The effort involved in implementing a new or modified coding system, once it has been developed, is relatively small because most of the work can be done within the existing dictionaries of noun and verb phrases. Once those dictionaries have been modified, even a long series of texts spanning multiple decades can then be recoded in a couple of minutes. This allows researchers to focus on maximizing the validity of a particular coding scheme because the automated coding process itself guarantees the reliability of the system.

Because of this flexibility, we anticipate that various projects within the event data coding community will continue to experiment independently with these coding methods

for a few more years before gradually converging on a set of common standards. This incremental approach contrasts to the top-down method used in the 1960s—and briefly attempted by the NSFs DDIR project in the early 1990s (Merritt, Muncaster and Zinnes 1993)—where attempts were made to find common coding conventions by getting a grant, holding a big meeting in a windowless conference room, and discussing coding standards based on assumptions about what the data might look like rather than through the slow and incremental experience of actual coding. These top-down efforts largely failed: the de facto standards have been set by the data sets themselves. Not coincidentally, this new process of experience-based standards would be similar to that found in the development of open-source software; with luck it will have similar success.

While we believe that the system described in this paper is a decided improvement over the previous work that we have been doing, there are still a few open issues. First, 9-character codes may not be sufficient in all instances, and more extended 12- and 15-character codes may be needed in some cases. Second, there may be opportunities to provide more general regional and ethnic coding. Finally, the TABARI coding software has not been adapted to handle hierarchical coding. Such a modification would simplify both the creation and maintenance of the actor dictionaries.

## 5.1   Extended codes

While the 9-character code string is sufficient for most applications, we have found a few situations, generally those involving ethnic conflict, where a 12-character code would be useful. These arise, for example, if one needed to distinguish multiple ethnically defined militia groups within a region within a state:

[Serbia] [Kosovo region] [ethnic Serb] [specific militia group]

[Serbia] [Kosovo region] [ethnic Albania] [specific militia group]

If one wished to additional identify specific individuals within those groupssomething that can be easily coded using automated methodsthese codes could be expanded to 15 characters.

On the one hand, this appears to encourage a proliferation of codes that will eventually result in a system of Byzantine complexity that would be impossible to use. Alternatively, as long as a consistent hierarchical structure is maintained, and as long as new three-character elements are created only when necessary, the amount of new complexity is actually limited. Contemporary statistical programs such as Stata and R have a full set of string-manipulation functions, as do text-processing languages such as perl, so the extended codes can be readily parsed into the relevant components if, as is typically the case, they are to be aggregated for statistical analysis. This approach of extracting information from a single hierarchically-organized actor field may be more efficient than trying to anticipate all of the possibly relevant codes and constructing a multi-field rectangular data structure to accommodate this.

## 5.2   Additional geographical and ethnic information

There appear to be at least two possible standard sources for cities, region and other sub-national codes. The ISO-3166-2 standard

`http://www.iso.org/iso/en/prods-services/iso3166ma/04background-on-iso-3166/iso3166-2.html`

provides 1- to 3-character alphanumeric codes for administrative regions within states. The number of codes varies depending upon the state but in many cases these are quite specificthe coding system for Italy has 20 regions and 103 provinces; Turkey has codes for 80 provinces.

A second possible source is the United Nations UN/LOCODE system

`http://www.unece.org/cefact/locode/service/main.htm`

that is used to track trade. It contains three-character codes for about 32,000 locations, primarily cities but also some ports and border-crossing points. This is generally quite thorough but like other UN coding systems, it does have some politically induced blind spots: Neither Gaza nor the West Bank are listed anywhere nor, interestingly, does the listing for Israel include Jerusalem.

The ISO also maintains a list of three-character codes for languages: ISO-639-2

`http://www.loc.gov/standards/iso639-2/englangn.html`

We are fully aware that language is not the same thing as ethnicity, and ISO-639-2 also provides some unnecessary codes we don't need: For example the Akkadians are unlikely to be causing trouble as the language died out 3,000 years ago. However, in many instances ethnicity is largely coterminous with languages, and consequently those codes would provide a solid basis for a list of ethnic group codes.

ISO-639-2 has all of the language codes in lower case, so if one established a norm of using lower-case letters for ethnic groups, one would always know that a code of the form "**AAAbbb**" so for example "**ISRpse**" is an ethnic group. In order to reduce the number of codes one needs to memorize, one could also establish the norm that a nationality that occurs as both a state and an ethnic group would have the state identification in upper case and the ethnicity in lower-case. So, for example, PSE is Palestine, pse refers to ethnic Palestinians; ALB is the Albanian state; alb refers to ethnic Albanians. We are currently not using this convention but may adopt it in the near future.

We've thought of three other possibilities for generating lists of ethnic groups; once these lists were established, a suitable set of mnemonic abbreviations could be assigned.

- process the CIA World Factbook to get the names of all of the ethnic groups recognized by the CIA

- Use idiosyncratic ethnographic sources. For example, the web site `http://www.2001pray.org/Africa.htm`, which like much of the ethnographic work

in North America is run by a missionary group, has a nice list of exactly 100 ethnic groups in Africa. However, this suspiciously round number suggest the origins of this list were in a two-column coding framework of some other data set, rather than a comprehensive survey of all ethnic groups in Africa...)

- `http://en.wikipedia.org/wiki/List_of_ethnic_groups` [accessed 20 March 2008] currently lists approximately 1,100 ethnic groups , and most of these entries have links to fairly extensive descriptions of the groups. While the flaws of Wikipedia are well-known (and well-publicized...), for the purposes of event coding this might be a nearly comprehensive source, as it seems unlikely that any ethnic group that is sufficiently politically-active to attract the attention of the international news media would not have at some point documented itself in Wikipedia.

These suggestions of supplemental sources are by no means exhaustive, and since we have not attempted to do global (or even continental) coding, we haven't done any research in depth on this matter, and better sources may exist, particularly in the anthropological literature. In fact, we would appreciate suggestions on this matter.

## 5.3  Extensions of TABARI to handle generalized codes

At the present time, the TABARI automated coding program (`http://web.ku.edu/keds/software.dir/tabari.html`) that we have been using to do automated coding has not been modified to reflect the changes in the actor coding scheme, so each code needs to be associated with an explicit noun phrase.

When the sub-state actor consists of a simple actor + agent pair—for example LIBERIAN_POLICE, NIGERIAN_POLICE, SENEGALESE_POLICE and so forth—it should be relatively straightforward to have TABARI concatenate the codes rather than having all of these separate entries in the actor dictionaries. This would have two advantages. First, it would reduce the size of the dictionaries and the effort involved in maintaining them. Second, whenever vocabulary was added for a new sub-state agent or role, it could immediately be applied to all of the states being coded, rather than requiring separate entries for each state. This modification is relatively straightforward.

A second, but more difficult, addition is to have TABARI fill in the primary actor code based on the geographical location of the event. A news wire story about a demonstration in Cairo may not explicitly refer to "Egyptian demonstrators" and "Egyptian police," assuming instead that the reader can infer that if the activity occurred in Egypt, the participants were Egyptian. To the extent that geographical location can be determined, and this might require separate dictionaries specifically for that task (for example, names of cities, regions, states and provinces), then the primary actor code could be added even when it is implicit.

As we write this paper in March 2008, a dominant news story involves the military conflict between Israel and various Islamic militant groups in Gaza. Israel is behaving as

a classical Westphalian state actor. However, its military operations are directed against a militarized non-state actor, and the territorial status of Gaza is utterly ambiguous. Various organizations within the United Nations are attempting to mediate the conflict and provide humanitarian aid. The United States is, characteristically, responding to the crisis as a Westphalian actor, but many of the European states are working through the EU. Street protests have erupted throughout the Arab and Islamic world in opposition to Israel and, more recently, in support of Hamas, and the outcome of this conflict is being watched closely as a possible model for other militarized movements, particularly in Iraq. This is the sort of world that we wish to study, and it is the sort of the world that the methods we are developing here are intended to code.

# References

[1] Azar, Edward E. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24:143-152.

[2] Azar, Edward E. 1982. *The Codebook of the Conflict and Peace Data Bank (COPDAB)*. College Park, MD: Center for International Development, University of Maryland.

[3] Azar, Edward E., and Thomas Sloan. 1975. *Dimensions of Interaction*. Pittsburgh: University Center for International Studies, University of Pittsburgh

[4] Bond, Doug; J. Craig Jenkins, Charles L. Taylor and Kurt Schock, 1997. "Mapping Mass Political Conflict and Civil Society: The Automated Development of Event Data." *Journal of Conflict Resolution* 41,4: 553-579.

[5] Bond, Doug, Joe Bond, Churl Oh, J. Craig Jenkins and Charles Lewis Taylor. 2003. "Integrated Data for Events Analysis (IDEA): An Event Typology for Automated Events Data Development" *Journal of Peace Research* 40,6: 733745

[6] Deborah J. Gerner, Philip A. Schrodt, Ömür Yilmaz and Rajaa Abu-Jabr. 2002. "Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions." Paper presented at the International Studies Association, New Orleans, March 2002 and American Political Science Association, Boston, August 2002.

[7] Deborah J. Gerner, Philip A. Schrodt, Ömür Yilmaz and Dennis Hermrick. 2005. "The CAMEO (Conflict and Mediation Event Observations) Actor Coding Framework." Paper presented at the American Political Science Association, Washington, September 2005.

[8] Deborah J. Gerner, Philip A. Schrodt and Ömür Yilmaz. 2007. *Conflict and Mediation Event Observations (CAMEO) Codebook*. Manuscript, http://web.ku.edu/keds/data.dir/cameo.html

[9] Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36: 369-385.

[10] International Studies Quarterly. 1983. "Symposium: Events Data Collections." *International Studies Quarterly* 27.

[11] Leng, Russell J. 1987. *Behavioral Correlates of War, 1816-1975.* (ICPSR 8606). Ann Arbor: Inter-University Consortium for Political and Social Research.

[12] King, Gary and Will Lowe. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* 57,3: 617-642.

[13] McClelland, Charles A. 1976. *World Event/Interaction Survey Codebook.* (ICPSR 5211). Ann Arbor: Inter-University Consortium for Political and Social Research.

[14] McClelland, Charles A. 1983. "Let the User Beware." *International Studies Quarterly* 27,2:169-177

[15] Merritt, R. L., R. G. Muncaster, and D. A. Zinnes, eds. 1993. *Theory and Management of International Event Data: DDIR Phase II*. Ann Arbor: University of Michigan Press.

[16] Schrodt, Philip A. and Deborah J. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-1992." *American Journal of Political Science* 38:825-854.

[17] Schrodt, Philip A. and Ömür Yilmaz. 2007. *CAMEO— Conflict and Mediation Event Observations—Codebook.* `http://web.ku.edu/keds/cameo.dir/CAMEO.CDB.09b5.pdf`

[18] Russett, Bruce M., J. David Singer and Melvin Small. 1968. "National Political Units in the Twentieth Century: A Standardized List." *American Political Science Review* 62, 3: 935-950

[19] Vincent, Jack E. 1979. *Project Theory: Interpretations and Policy Relevance*. Washington: University Press of America.