

The Composition of Religious and Ethnic Groups (CREG) Project

October 2012

Cline Center for Democracy
University of Illinois at Urbana-Champaign



ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

The Composition of Religious and Ethnic Groups (CREG) Project

ABSTRACT

The CREG project is a multi-faceted effort to compile information on the nature and depth of socio-cultural cleavages for 165 countries in the postwar era. This white paper provides some basic and preliminary information on the data collection phase of this project, which has two components. One describes the assembly of data on the relative size of the main ethnic and religious groups in each country. The second component describes the objectives and design of a web-based survey that will use an “expert-sourcing” approach to obtain information on the types of principal socio-cultural tensions in a country, the nature and depth of socio-cultural cleavage, and some locational information on the sub-national distribution of a country’s ethnic and religious groups.

CONTRIBUTORS

Peter F. Nardulli
Cara J. Wong
Ajay Singh
Buddy Peyton
Joseph Bajjalieh

Table of Contents

Introduction	2
Gauging the Socio-cultural Composition of Countries	2
The Sources.....	3
Data Assembly Procedures	3
The Annual Projections.....	5
Gauging the Attributes of Socio-cultural Groups	8
References	11
Appendix I	12
Appendix II	15
Sensory-based Distinctions.....	16
<i>Physical Appearances</i>	16
<i>Language</i>	16
<i>Cultural Styles</i>	16
<i>Leisure Activities</i>	17
Ascribed Differences Embedded in Group-specific Conventional Wisdoms.....	17
Group Affiliations.....	18
<i>Religious</i>	18
<i>Socio-economic Status</i>	18
<i>Political</i>	18
Domain-specific Relationships	19
<i>Social</i>	19
<i>Economic</i>	20
<i>Legal</i>	20
<i>Political</i>	21

The Composition of Religious and Ethnic Groups (CREG) Project

Introduction

The Composition of Religious and Ethnic Groups (CREG) project is part of the Cline Center for Democracy's Societal Infrastructures and Development (SID) project. Its objective is to fill an important information void in cross-national studies by creating a set of time-varying measures that gauge the nature and depth of socio-cultural cleavages. Its focus is on the 165 countries included in the SID project (all countries with a population above 500,000 in 2004) during the post-WWII era. The data-gathering phase has two main components: one deals with the socio-cultural composition of countries; the other focuses on the attributes of a country's main socio-cultural groups. The compositional component involves identifying, collecting and analyzing available data on the relative sizes of the ethnic and religious groups in a country. Its aim is to create annual, country-specific projections for the various groups. The attributional component involves collecting and analyzing new data using a web-based survey of country and regional experts. Its aim is to generate a set of measures that capture the nature and depth of socio-cultural cleavages for the main ethnic and religious groups identified in the compositional component. The following sections provide details on each component.

Gauging the Socio-cultural Composition of Countries

Because of the daunting challenges involved in generating population projections for a large number of groups in 165 countries, we conducted a pretest to determine the feasibility of such an initiative. The aim of the pretest was to ascertain whether available sources contained data for most or all of the main socio-cultural groups. We also wanted some sense of whether: (1) the data from the available sources yielded consistent estimations of population sizes and (2) there was enough compatible data spread over a long enough time span to generate projections for the postwar era. To conduct the pretest we began by examining the country-specific ethnic group configurations identified in the *Geo-referencing of Ethnic Groups Project* (GREG).¹ Secondly we selected three countries from different parts of the world that data from the GREG project suggested were ethnically diverse: Afghanistan, Indonesia, and Sudan. We then consulted three sources that a literature review suggested were the most widely used sources for information on ethnic and religious data. They are: the Britannica Book of the Year (BBOY), the CIA World Factbook (CIA-WF) and the World Almanac Book of Facts (WABF).

A review of these sources revealed that data for virtually all of the ethnic groups included in the GREG data for these three countries were available in one or all of the sources. We then collected and integrated the available data on ethnic and religious groups from all three sources and displayed them on a set of graphs with year as the x-axis. These graphs demonstrated that the available data on country-specific groups were consistent enough across sources, and sufficiently distributed across time, to generate projections of relative population sizes for the postwar era. The pretest also made it clear that generating well-grounded projections efficiently would involve a major effort requiring: (1) coordinated efforts by two research teams (a data collection/

¹ For more information on GREG see: <http://www.icr.ethz.ch/data/other/greg>.

assimilation team and a data analysis team); (2) the use of supplemental sources for specific country-groups and years; and (3) carefully designed and implemented statistical procedures.

The Sources

We used the three sources listed above as our basic source of information for ethnic and religious group populations.² The BBOY is a compendium of the major events that occurred during the previous year, organized by topic and/or country. The types of information available in BBOY, and the manner it is reported, changed over time. The earlier editions (1945 through 1985) organized their country-specific information with alphabetically arranged country profiles. These earlier issues contained sparse information on demographics. In editions appearing after 1985, BBOY created a section called World Data, containing all country profiles organized in alphabetical order. It also included a section on comparative national statistics in which it reported various pieces of demographic and economic data, including population data.

The CIA World Factbook contains an array of country-specific data on economics, demographics, geography and national defense. The CIA's early efforts in compiling these data culminated in the 1962 publication of the National Basic Intelligence Factbook (NBIF), a classified document. The first declassified version appeared in 1971, and by 1981, the CIA changed the name of the NBIF to the World Factbook. The World Almanac Book of Facts is a reference book that began publishing in 1868. It is similar in design to the Britannica Book of the Year in that it contains information on the major events that occurred within the past year.

While all three of these sources contains data on the relative size of ethnic and religious groups at some point in their publication run, the availability and format of the data varies across sources and over time. For example, the pre-1980 editions of BBOY and WABF contain sparse statistical data on ethnic and religious groups, while all editions of the CIA-WF report data on them. When population data are reported they are usually reported as a proportion of the entire population. Moreover, the available data on group sized are normally reported in textual discussions of a country, not in a table. The main exceptions here are CIA-WF, and post-1985 editions of BBOY and WABF. Perhaps the most important point concerning the available population data on these groups is that most of it is not accompanied by a year citation. The only exception is the post-1985 editions of the BBOY. Another point worth noting is that sequential publications frequently cite the same figure for a particular group (e.g., Russians are reported as composing 40% of the population in Estonia for 1992 through 2002). Finally, it should be noted that, in rare instances, different editions of a source provide inconsistent data for the same year (e.g., the 1989 edition of BBOY reports that Pashtuns in Afghanistan constitute 58% of the population in 1985 while its 1992 edition reports the 1985 figures as 65%)

Data Assembly Procedures

Issues related to format of the available data led to a two-tiered structure for collecting and integrating them. The first tier consisted of five data collectors; at the second tier were two data

² With respect to the BBOY, we obtained all editions between 1945 and 2000; for the post-2000 era, however, we only had available the 2002, 2005 and 2010 editions. The CIA-WF existed in publically accessible form between the years 1981 and 2012, so it was of limited use in providing early data. We obtained all editions of WABF between 1945 and 2012.

integrators. The data collectors were assigned a set of countries (three at a time) and tasked to record all the relevant information from the three sources noted above – including inconsistencies and duplicated entries – into a spreadsheet. These data entries were accompanied by information on the source and year of the edition from which they were derived – or the year to which the data were ascribed, if available. Each spreadsheet also contained a “Notes” worksheet that the data collectors used to enter information that would potentially be useful to the data integrators.

After the data collectors recorded all of the available information, they forwarded the spreadsheet to the integrators, who always processed the data jointly. Their task was to check and reconcile the data provided by the collectors and record it in a new spreadsheet in which the data are organized by group, year and source. The integrators performed a number of checks before the data were entered into the final spreadsheet. The first had to do with the consistency of group names. The same group is often referenced using different names across sources and over time (e.g., Pashtuns, Pathans, Pushtuns, Pashas; Hungarians, Magyars). Once it once confirmed that the reference was to the same group, a single term was selected and used for all entries. The second check involved data outliers (e.g., a group that is reported as constituting 25% of the population in one year and 35% in the next) by re-examining the source of the information. Corrections were made if the integrators’ suspicions were confirmed.

The third check was for data inconsistencies (e.g., different editions of the same source reports a group as constituting 18% of the population and 26% of the population in 1968). These inconsistencies were checked by re-examining the source. If the figures were confirmed the inconsistency was resolved as follows. If one of the figures is cited in only one edition (1979) while another is cited a series of editions (1976 through 1981), then the data most frequently cited was used. If neither figure was cited in a series of editions, the most recent figure was used. Thus, if one figure for 1990 was cited in a 1992 edition of a source, but a 1996 edition reported a different figure for 1990, the figure from the 1996 edition was used. The next check was to address situations involving the repetition of the same information across a series of years (e.g., instances in which Uzbeks were reported as constituting 8% of the population in Afghanistan for the period from 1962 through 1975). Because these situations were almost certainly the result of the same census data, including them would skew the calculations used to make the annual projections. Thus, only the data for the first year in the series was entered into the spreadsheet.

The last task performed by the integrators was to assign a year to the entries that did not have one associated with them. They assigned a date by determining the year of the earliest edition of the source in which the data were reported; the year of the earliest edition was used as a proxy for the missing date. For example, if Roma are first reported as constituting 6% of the Bulgarian population in the 1950 edition of WABF – and no year is associated with that figure – then 1950 was assigned as the date for that entry. If new information is reported on the Roma in the 1965 edition of BBOY (e.g., they constituted 9% of the Bulgarian population), and no year was associated with the new figure, then the integrators assigned 1965 as the date for that entry.

The Annual Projections

After the raw data were integrated into a spreadsheet, the integrators transmitted the spreadsheets to the data analysis team. Generating annual projections required a number of steps. First, the data analysts constructed a dataset for each country from the integrated spreadsheets. As the assembled data were gathered from multiple sources, there are situations where there is more than a single reported value for one or more groups in a given year. Second, they generated a set of preliminary projections using a multi-level mixed model with fixed group effects and random year effects. The dependent variable was group i 's logged portion of the population in a given year j ($\text{POP}\%_{ij}$).³ The explanatory variables were a linear estimator (INDEX)⁴, a group dummy variable (GRP_i), an interaction between the group dummy variable and the linear estimator ($\text{GRP}_i*\text{INDEX}$), a random disturbance term (v_j) associated with each year, and error term (ϵ_{ij}).⁵ The equation estimated was:

$$\text{POP}\%_{ij} = \alpha + \beta_1\text{INDEX} + \beta_{11}\text{GRP}_i + \beta_{12}\text{GRP}_i*\text{INDEX} + \dots + \beta_{i1}\text{GRP}_i + \beta_{i2}\text{GRP}_i*\text{INDEX} + v_j + \epsilon_{ij}$$

This approach allows each group to have a unique intercept and slope.

The data analysts used the data from the preliminary predictions to produce two graphical displays. Examples of these displays are depicted in Figure 1 and Figure 2, which report the results for ethnic groups in Afghanistan. Figure 1 represents a sum of the predicted $\text{POP}\%_{ij}$ for all ethnic groups; under ideal circumstances these annual sums would equal 100. Figure 2 is an overlay between constrained and unconstrained estimates where the unconstrained estimate is the predicted value of $\text{POP}\%_{ij}$ and the constrained estimate is the ratio of each group to the sum of estimates, which forces the values to sum to 100.⁶ In cases where the sum of the predicted values of $\text{POP}\%_{ij}$ is close to 100, the two estimates are essentially identical.

The displays illustrated in Figure 1 provided a quick visible check on the adequacy of the model projections. If the annual sums revealed large and consistent deviations from 100, then it was clear that further analyses were required. The displays illustrated in Figure 2 provided insights into what needed to be done to improve the projections. The three most significant problems they revealed pertained to deviant data points (i.e., outliers), missing groups, and troublesome data gaps (usually appearing at the beginning and/or end of the time-series).

³ Preliminary tests revealed that, when estimating values for relatively small groups, the linear trend predictor sometimes produced negative estimated values. To eliminate this problem, the $\text{POP}\%_{ij}$ variables were logged prior to estimation. Once the estimates for each country were made the predicted values transformed back to their original scale, producing an estimated annual population proportion for each group in a country.

⁴ To create the INDEX variable we used a simple count of the years between 1945 and 2013 (1946=1; 1947=2, 1948=3, etc.). The INDEX variable was used to estimate each group's proportion of the population for every year within that period. This approach was adopted because, while significant secular changes can occur over long stretches of time, relative groups sizes rarely vary much from year to year. Where non-linear data trends were evident, non-linear terms were estimated in the model.

⁵ For a country with n groups, $n-1$ group dummy variables and interactions were included in the model specification.

⁶ Constrained estimate = $(\text{POP}\%_{ij} / \text{POP}\%_{ij}\text{SUM}) * 100$

Figure 1
Annual Sum of Estimated Ethnic Group Percentages for Afghanistan

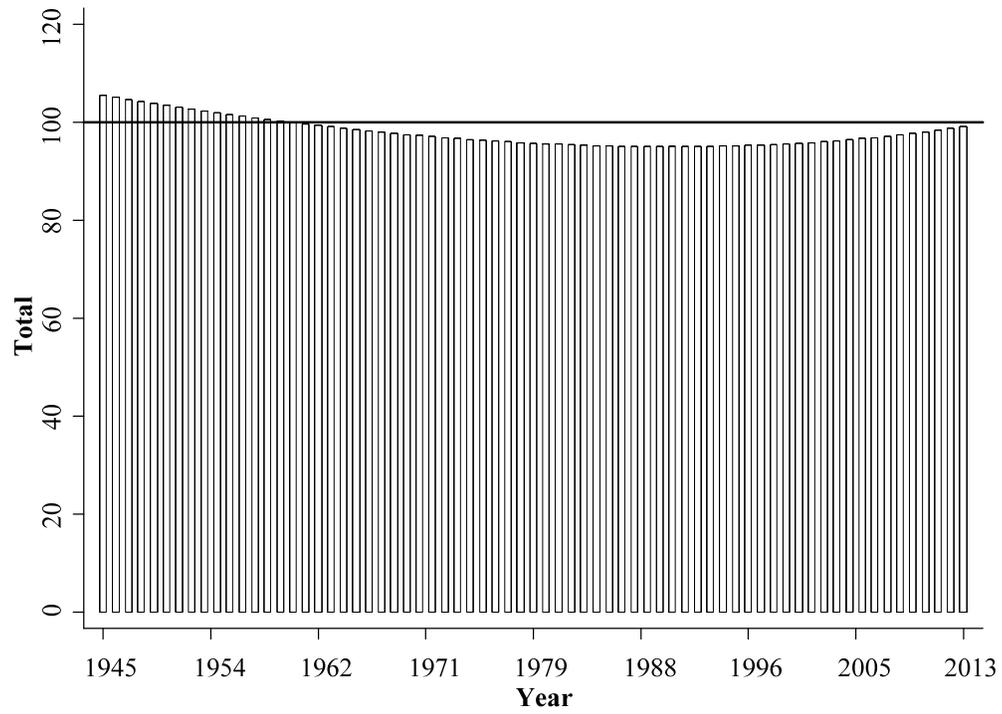
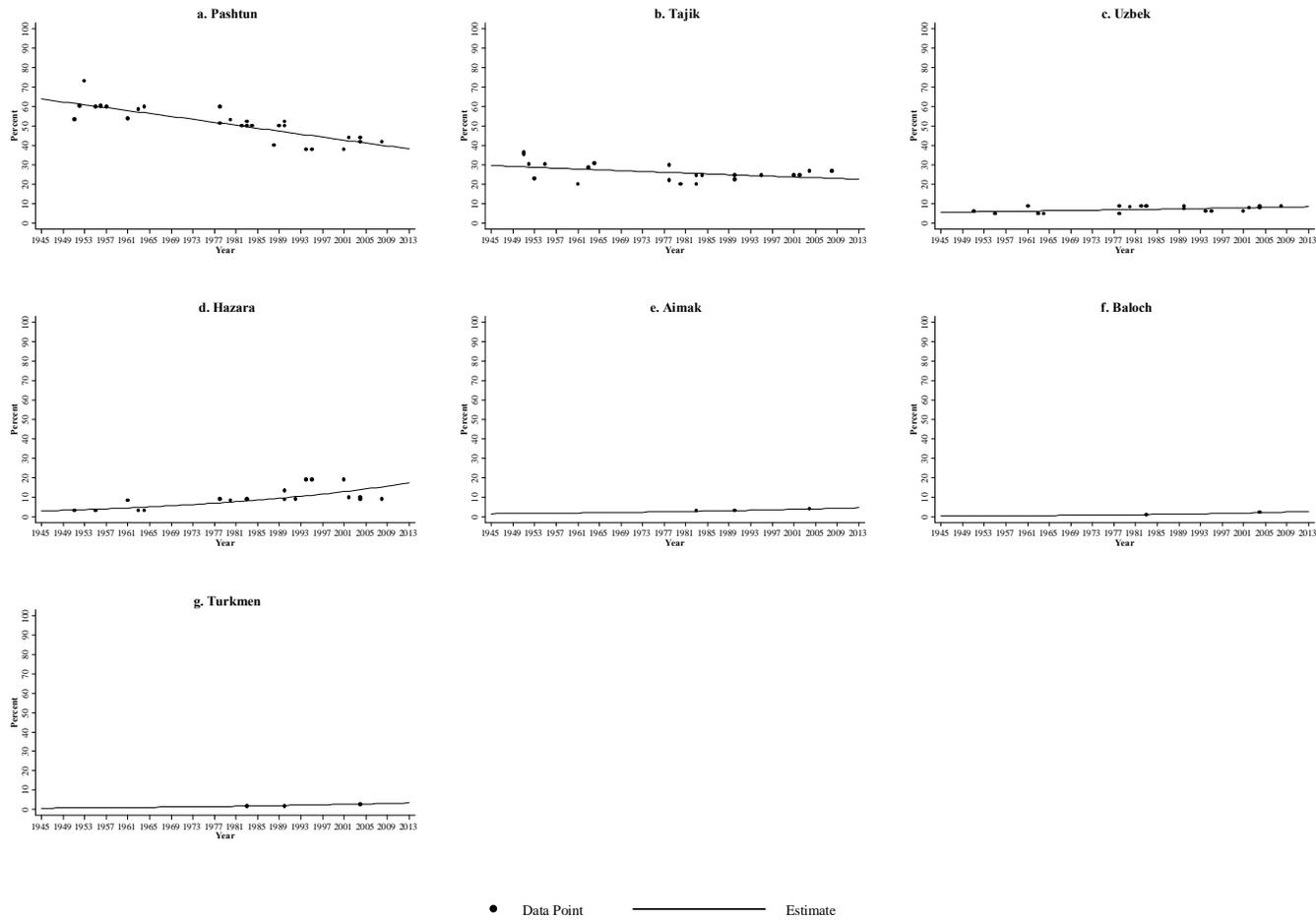


Figure 2
Overlay between Data Points and Unconstrained Estimates for Afghanistan Ethnic Groups



Where outliers were apparent additional projections were produced with and without them in the model. When the projections improved markedly, the outliers were eliminated and recorded in a dummy variable depicting outliers. Outliers account for 1.4% of the integrated data. In some countries, the sum of the predicted values of **POP%_{ij}** fell well short of 100 because a number of small groups were not included in the raw data. In some cases this created a situation in which the available data fit the unconstrained projections very well but the constrained projections deviated from the unconstrained ones. In these situations, when the existence of a large number of small groups was confirmed,⁷ an “other group” variable was created. The values assigned this variable in a given year was the difference between the sum of the unconstrained predicted values for all other groups and 100.

When data gaps were identified in displays such as Figure 2 they were recorded by country, group and time frame (e.g., the Indian population in Fiji for the 1950’s) and incorporated into a list that was transmitted to the data collection team. The transmittal of these lists initiated the second phase of the data collection effort. It involved searching library and online resources for supplemental sources containing group- and date-specific information on the relative size of ethnic and religious groups. To conduct these targeted searches the data collectors consulted librarians as well as library and on-line search engines. The sources identified often referenced other sources, which were then pursued. The search for additional sources included global, regional and country specific sources. While hundreds of potential resources were identified, only a fraction contained useful/needed information. The sources used are listed in Appendix I.

Once the additional data was added into an augmented spreadsheet (using the same procedures outlined above), the spreadsheets were returned to the data analysis team. At that point, they produced another set of projections using the procedures discussed above.⁸ In the vast majority of cases supplementary data was found and the final projections were markedly improved, with the sums of the expected values of **POP%_{ij}** being very close to 100. The median of the minimum and maximum values for these sums are 98 and 102, respectively; the median gap between the minimum and maximum values was 4.⁹ Illustrations of the final sets of projections can be found at the following site:

<http://www.clinecenter.illinois.edu/research/CREG/Examples/EthnicGroups.html>.

Gauging the Attributes of Socio-cultural Groups

As the illustrative on-line graphs demonstrate, our projections do a good job in capturing the relative sizes of the different socio-cultural groups in a country, as well as how those proportions have changed over time. What the projections cannot do, however, is provide insights into such

⁷ Sources such as Ethnic Groups World Wide, Book of the Peoples of the World, or The Encyclopedia of the Peoples of the World were used to identify these cases.

⁸ In most cases the projections for **POP%_{ij}** were calculated for the period from 1945 through 2013. In cases where a country became independent after 1945, the projections were for every year from independence through 2013. For example, Mexico has estimates for each group from 1945 to 2013; Guinea-Bissau, which gained independence in 1974, has estimates from 1974 through 2013.

⁹ The minimum sum had a mean of 97.6, the maximum a mean of 101.6, and the mean gap was 4.1.

things as: what differentiates the various groups from one another (physical features, language, cultural styles, ascribed characteristics pertaining to such things as intelligence, work ethic, moral turpitude, etc.); how different the groups are from one another with respect to differentiating attributes; how those inter-group differences have changed over time; or how the constellation of group differences varies across countries at a given point in time. Moreover, the compositional data does not tell us whether country-specific contextual factors have operated to ameliorate/aggravate inter-group differences (e.g., crosscutting organizational memberships; dyadic relationships within social, economic, legal and political domains that reflect power differentials). These mediating factors can affect the level of socio-cultural animosities in a society as well as how they are manifested in civil society.

Because of the importance of some of the issues raised above, and the inability of the compositional data to shed light on them, a survey-based approach was developed to acquire more information on the attributes of the main socio-cultural groups in a country. To acquire this information an “expert sourcing” approach is being pursued in conjunction with a survey that is administered by a sophisticated web-based survey engine. The website survey allows access, on a country-specific basis, for carefully selected experts. These experts will be asked to respond to extensively pre-tested questions relating to differences between country-specific groups for all major groups for which demographic data are available. Moreover, the survey engine is designed to gauge those socio-cultural distances at different points in time across the postwar era, where applicable. Finally, the survey will capture relevant locational data for the various groups. Collecting these attributional data will complement the compositional data and enhance its value.

While the survey is in an early stage of development, its broad outlines have been established. It will be composed of four main components. The first component deals with sensory-based and ascribed differences between groups. The second concerns country-specific factors that ameliorate/aggravate inter-group differences. The other sections relate to group-specific geographic concentrations within the country and the existence of politically meaningful diasporas outside the country. An outline of the first two sections, which are the organizationally most complex, is provided below.

Section One: Group-based Differences

1. Sensory-based (e.g., sight, sound, taste) matters that gauge differences in:
 - a. Physical appearances
 - i. Skin color, facial attributes, hair, physical stature
 - b. Linguistic differences
 - i. Language, use of words/phrases, accents
 - c. Cultural styles
 - i. Distinctive clothing, customary accoutrements (necklaces, jewelry head gear, etc.), mutable styles (hair length, facial hair, piercings, tattoos, etc.)
 - d. Leisure activities
 - i. Food, music
2. Ascribed differences derived from group-based conventional wisdoms pertaining to such matters as:
 - a. Intelligence
 - b. Work habits
 - c. Personal hygiene
 - d. Alcohol consumption
 - e. Drug abuse
 - f. Sexual mores
 - g. Religiosity

Section Two: Country-specific Mediating Factors

1. Group affiliations that could either bridge or aggregate group-based difference based on observable differences or conventional wisdoms
 - a. Religious affiliations
 - b. Social class
 - c. Political allegiances
2. Domain-specific matters that shed light on the status of inter-group relationships
 - a. Social
 - b. Economic
 - c. Legal
 - d. Political

A very rough and preliminary draft of the web-based survey is presented in Appendix II. It illustrates the approach to be used to for capturing information on ethnic groups using Pashtuns and Tajiks in Afghanistan.

References

Davis, Wade, K D. Harrison, and Catherine H. Howell. *Book of Peoples of the World: A Guide to Cultures*. Washington, D.C: National Geographic, 2007. Print.

Gonen, Amiram. *The Encyclopedia of the Peoples of the World*. New York: H. Holt, 1993. Print.

Levinson, David. *Ethnic Groups Worldwide: A Ready Reference Handbook*. Phoenix, Ariz: Oryx Press, 1998. Print.

Olson, James S. *The Peoples of Africa: An Ethnohistorical Dictionary*. Westport, Conn: Greenwood Press, 1996. Print.

Weidmann, Niels B, and Ketil R. L.-E. C. Jan. "Representing Ethnic Groups in Space: A New Dataset." *Journal of Peace Research* (trykt Utg.). 47 (n.d.). Print.

Appendix I Supplemental Sources Used to Assemble Ethnic Groups Data Sets

- Abshire, D. M., Samuels, M. A., Georgetown University. Center for, S., & International, S. (1969). *Portuguese Africa; a handbook*. New York: Published in cooperation with the Center for Strategic and International Studies, Georgetown University, by Praeger.
- Adeney, K. (2007). *Federalism and ethnic conflict regulation in India and Pakistan*. New York: Palgrave Macmillan.
- Aleaz, B., Ghosh, L., & Dutta, A. K. (2003). *Ethnicity, nations and minorities: the South Asian scenario*. Delhi: Indian Association for Asian & Pacific Studies in association with Manak Pub.
- Appiah, A., & Gates, H. L. (2010). *Encyclopedia of Africa*. Oxford; New York: Oxford University Press.
- Baer, G. (1964). *Population and society in the Arab east*. New York: Praeger.
- Banuazizi, A., Weiner, M., Joint Committee on the, N., Middle, E., & Joint Committee on South, A. (1986). *The State, religion, and ethnic politics : Afghanistan, Iran, and Pakistan*, [Syracuse, N.Y.].
- Berger, M. (1962). *The Arab world today*. Garden City, N.Y.: Doubleday.
- Caldarola, C. (1982). *Religions and societies, Asia and the Middle East*. Berlin; New York: Mouton.
- Clarke, C. G. (1991). *Society and politics in the Caribbean*. New York: St. Martin's Press.
- Coleman, J. S., & Rosberg, C. G. (1964). *Political parties and national integration in tropical Africa*. Berkeley: University of California Press.
- Curtis, G. E., & Library of Congress. Federal Research, D. (1997). *Kazakstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan : country studies*. Washington, D.C.: The Division : For sale by the Supt. of Docs., U.S. G.P.O.
- De Silva, K. M. (1988). *Ethnic conflict in Buddhist societies : Sri Lanka, Thailand, and Burma*. London; Boulder, Colo.: Pinter ; Westview Press.
- Dzidzienyo, A., & Casal, L. (1979). *The position of blacks in Brazilian and Cuban society*. [London]: Minority Rights Group.
- Eberhardt, P. (2003). *Ethnic groups and population changes in twentieth-century Central-Eastern Europe : history, data, and analysis*. Armonk, N.Y.: M.E. Sharpe.

- Economist. (1981). *The World in figures*. London: Economist Newspaper.
- Emigh, R. J., & Szelényi, I. (2001). *Poverty, ethnicity, and gender in Eastern Europe during the market transition*. Westport, Conn.: Praeger.
- Encyclopaedia Britannica, i. (1938). Britannica book of the year. *Britannica book of the year*.
- Europa Publications, L. (1971). Africa South of the Sahara. *Africa South of the Sahara*.
- George, J. A. (2009). *The politics of ethnic separatism in Russia and Georgia*. New York: Palgrave Macmillan.
- Hawkins, F. (1989). *Critical years in immigration : Canada and Australia compared*. Kingston [Ont.]: McGill-Queen's University Press.
- Information Please, L. L. C. (1998). Infoplease.
- Israel. Lishkat, h.-i., & Israel. Misrad ha-huts. ha-Lishkah, l.-h. (1952). Facts about Israel. *Facts about Israel*.
- Kaplan, I., & American University Foreign Area, S. (1976). *Area handbook for Sierra Leone*. Washington: for sale by the Supt. of Docs., U.S. Govt. Print. Off.
- Kurian, G. T. (1992). *Encyclopedia of the Third World*. New York: Facts on File.
- Mansfield, P., & Royal Institute of International Affairs. Information, D. (1973). *The Middle East; a political and economic survey*. London; New York: Oxford University Press.
- Mbanaso, M. U., & Korieh, C. J. (2010). *Minorities and the state in Africa*. Amherst, NY: Cambria Press.
- Misra, S. S. (1995). *Ethnic conflict and security crisis in Sri Lanka*. Delhi: Kalinga Publications.
- Nelson, H. D., & American University . Foreign Area, S. (1975). *Area handbook for Southern Rhodesia*. Washington: For sale by the Supt. of Docs., U.S. Govt. Print. Off.
- Sanders, A. J. K. (1968). *The People's Republic of Mongolia: a general reference guide*. London; New York: Oxford U.P.
- Siegel, J. S. (1958). *The population of Hungary*. Washington: United States G.P.O.
- The time almanac. (1998). *The time almanac*.

United Nations. Statistical, O., United Nations. Statistical, D., United Nations. Dept. of Social, A., United Nations. Dept. of Economic, A., United Nations. Dept. of, E., Social, A., et al. (1948). Demographic yearbook. *Annuaire démographique. Demographic yearbook. Annuaire démographique.*

United States. Central Intelligence, A. (2008). The CIA world factbook. *The CIA world factbook.*

Vreeland, N., Clare, K. G., & American University . Foreign Area, S. (1975). *Area handbook for South Korea.* Washington: For sale by the Supt. of Docs., U.S. Govt. Print. Off.

Weidmann, N. B., Jan Ketil, R., & Lars-Erik, C. (2010). Representing ethnic groups in space : a new dataset. *Journal of peace research (trykt utg.).* 47.

Whitaker's almanac 1974. (1973). London: J. Whitaker & Sons.

The World almanac and book of facts. (1923). *The World almanac and book of facts.*

Appendix II
Preliminary Sketch of CREG Survey: Pashtun/Tajik Comparisons for Afghanistan

The census data for Afghanistan during the Post-World War II era indicate that the major non-religious socio-cultural groups are as follows:

Major groups (constituted at least X% of the country's population at some point in the post-war era):

1. Pashtun
2. Tajik
3. Uzbek
4. Hazara

Minor groups (constituted at least X% of the country's population at some point in the post-war era):

1. Aimak
2. Turkmen
3. Baloch

Are there any other politically significant, non-religious groups in Afghanistan?

Yes

No

If yes, please specify the group and the decade when they became politically significant

Group

Decade

Sensory-based Distinctions

Physical Appearances

If a group of **Pashtuns** and **Tajiks** dressed in identical attire were standing silently on a street corner in **Afghanistan**, how easy would it be to differentiate between them in terms of:

Skin color (Sliding bar from Very difficult to Very easy)

Facial attributes (Sliding bar from Very difficult to Very easy)

Hair (Sliding bar from Very difficult to Very easy)

Stature (Sliding bar from Very difficult to Very easy)

Language

If a group of **Pashtuns** and **Tajiks** day laborers were assigned to work on a construction project in **Afghanistan** in 1950, how difficult would it be for them to communicate with one another based on their:

Native language (Sliding bar from Very difficult to Very easy)

Word/phrase usage (Sliding bar from Very difficult to Very easy)

Accents (Sliding bar from Very difficult to Very easy)

Would your answers change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Cultural Styles

If a group of **Pashtuns** and **Tajiks** dressed in traditional attire were standing silently on a street corner in **Afghanistan** in 1950, how easy would it be to differentiate between them in terms of their:

Clothing (Sliding bar from Very difficult to Very easy)

Accoutrements (jewelry, head gear, etc.) (Sliding bar from Very difficult to Very easy)

Appearance alterations (facial hair, hair length, cosmetics, body piercing, tattoos, etc.) (Sliding bar from Very difficult to Very easy)

Would your answers change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Leisure Activities

If a group of **Pashtuns** and **Tajiks** were planning a joint event in **Afghanistan** in 1950, how easy would it be for them to agree on the type of:

Food to be served (Sliding bar from Very difficult to Very easy)

Music to be played (Sliding bar from Very difficult to Very easy)

Would your answers change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Ascribed Differences Embedded in Group-specific Conventional Wisdoms

In 1950 what image would **Pashtun** conventional wisdom convey about differences between the average **Pashtun** and the average **Tajik** in terms of their

Intelligence (Sliding bar from Far less intelligent to Far more intelligent);

Not part of the conventional wisdom

Work habits (Sliding bar from Far more industrious to Far less industrious);

Not part of the conventional wisdom

Personal hygiene (Sliding bar from Far less concerned with hygiene to Far more concerned with hygiene);

Not part of the conventional wisdom

Alcohol consumption (Sliding bar from Far less likely to overindulge to Far more likely to overindulge);

Not part of the conventional wisdom

Drug abuse (Sliding bar from Far less likely to abuse drugs to Far more likely to abuse drugs);

Not part of the conventional wisdom

Sexual mores (Sliding bar from Far less promiscuous to Far more promiscuous);

Not part of the conventional wisdom

Religious fervor (Sliding bar from Far less religious to Far more religious); Not part of the conventional wisdom

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Group Affiliations

Religious

Were there significant differences in the religious affiliation of the various groups mentioned above in 1950?

If yes, please use the embedded tool to estimate the affiliation of each group with the following religious groups, on a scale of 0 to 100

What proportion of **Pashtuns** would you estimate to be **Sunni Muslims**?

What proportion of **Pashtuns** would you estimate to be **Shia Muslims**?

What proportion of **Pashtuns** would you estimate to be **Hindus**?

What proportion of **Pashtuns** would you estimate to be **Zoroastrians**?

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Socio-economic Status

Were there significant differences in the socio-economic status of the various groups mentioned above in 1950?

If yes, please indicate the relative differences using the embedded tool to depict the relative standing of the different groups in 1950

Higher socio-economic status

Medium socio-economic status

Lower socio-economic status

Pashtun Tajik Uzbek Hazara Aimak Turkmen Baloch

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

1) yes

2) no

[if yes, repeat sequence for new year]

Political

Were there significant differences in the political allegiances of the various groups mentioned above in the 2000's?

If yes, please use the embedded tool to estimate the allegiance of each group with the following national political parties, on a scale of 0 to 100

What proportion of **Pashtuns** were aligned with the **National Liberation Front**

What proportion of **Pashtuns** were aligned with the **National Congress Party of Afghanistan**

What proportion of **Pashtuns** were aligned with the **National Solidarity Movement of Afghanistan**

What proportion of **Pashtuns** were aligned with the **Afghanistan Independence Party**

Domain-specific Relationships

Social

In 1950, how likely would it be for a **Pashtun** to have a neighbor who is a **Tajik**?

- 1) Very likely
- 2) Somewhat likely
- 3) Not very likely at all

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

- 1) yes
- 2) no

[if yes, repeat sequence for new year]

In 1950, how likely would it be for a **Pashtun** to be friends with a **Tajik**?

- 1) Very likely
- 2) Somewhat likely
- 3) Not very likely at all

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

- 1) yes
- 2) no

[if yes, repeat sequence for new year]

In 1950, how difficult would it be for a **Pashtun** to marry a **Tajik**?

- 1) Very difficult
- 2) Somewhat difficult
- 3) Not very difficult at all

[If R chooses "very" or "somewhat difficult"]

Would the difficulty be a result of:

- 1) Anti-miscegenation laws
- 2) Social norms
- 3) Religious impediments

Respondent can select more than one.

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

- 1) yes
- 2) no

[if yes, repeat question for new year]

Economic

Imagine a **Tajik** is applying for a job; he is an upstanding citizen and is qualified for the job. If the potential new employer is a **Pashtun**, how likely is it that the **Tajik** will be treated fairly in the hiring process in 1950?

- 1) Very likely
- 2) Somewhat likely
- 3) Not very likely at all

How likely would this scenario have been?

- 1) Very likely
- 2) Very unlikely given power dynamics in the country
- 3) Very unlikely given that (Group A) and (Group B) had very little contact given where they lived.

Would your answer change if the year were (1960, 1970, 1980, 1990, 2000, 2010)?

- 1) yes
- 2) no

[if yes, repeat question for next year]

Legal

Imagine a **Tajik** is walking down a street, when he is stopped by the local law enforcement in 1950. He is innocent of any wrongdoing, but the police had received a call about a suspicious person. If the police officer who stopped the man is a **Pashtun**, how likely is it that the **Tajik** will be treated fairly?

- 1) Very likely
- 2) Somewhat likely
- 3) Not very likely at all

How likely would this scenario have been?

- 1) Very likely
- 2) Very unlikely given power dynamics in the country
- 3) Very unlikely given that (Group A) and (Group B) had very little contact given where they lived.

Would your answer change if the year were (year)?

- 1) yes
- 2) no

[if yes, repeat vignette for new year]

Political

Imagine a **Pashtun** is the head of a government agency charged with the responsibility for road construction and maintenance in **1950**. How likely is it that **Tajik** areas will receive equitable allocations of funding?

- 1) Very likely
- 2) Somewhat likely
- 3) Not very likely at all

How likely would this scenario have been?

- 1) Very likely
- 2) Very unlikely given power dynamics in the country
- 3) Very unlikely given that **Tajiks** did not live in segregated areas.

Would your answer change if the year were (year)?

- 1) yes
- 2) no

[if yes, repeat vignette for new year]