BROWN

Department of
Economics

January 10th, 2021

To:
Those Interested in Exploring the Catalogue of Oral Traditions

RE: Replication Materials for MS 30574

Title: Folklore

Authors: Stelios Michalopoulos, Melanie Meng Xue

In Section 1 of this note, we describe the files on oral traditions across (i) 958 groups in Berezkin, (ii) 1,245 groups in the Ethnographic Atlas (EA), and (iii) 199 countries. In Section 2, we detail how to regenerate the Tables and Figures that appear in the Main text and the Appendix of the manuscript.

Section 1

Original Files on Folklore **(folder "Original Files")**

1) Bereskin's original motif catalogue of 2,564 motifs. Each motif comes with a title and a short description. we provide the title and the description (i) in Russian, (ii) in English translated by Berezkin himself, and (iii) in English translated by Google API translation services. The catalogue version is current as of October 2019. Typos, gaps, and repetitions reflect those in the original catalogue.

File: "Motif_Master.dta"

Brown University, 64 Waterman Street, Providence, RI 02912. TEL: 401 863-3836.Email: smichalo@brown.edu

2) Incidence of the 2,564 motifs across:

   a. 958 groups in Berezkin **File: "Motifs_Berezkin_groups.dta"**

   b. 1,245 groups in the EA **File: "Motifs_EA_groups.dta"**

   c. 199 countries **File: "Motifs_Countries.dta"**

3) Folklore-based measures of:

   (i) trust (proxied by how tricksters' antisocial behavior is portrayed),

   (ii) risk-taking (captured in competitions and challenges being harmful or not), and

   (iii) gender norms (reflected in the various stereotypes of males and females characters).

   Across:

   a. 958 societies in Berezkin's **File: "MTurk_Berezkin_groups.dta"**

   b. 1,245 groups in the EA **File: "MTurk_EA_groups.dta"**

   c. 199 countries **File: "MTurk_countries.dta"**

For completeness, we include the original classifications by MTurks in the following files:

Brown University, 64 Waterman Street, Providence, RI 02912. TEL: 401 863-3836.Email: smichalo@brown.edu

- **File: "male_female_motifs_classifications.dta"** contains the MTurks classification of the 1,073 motifs into male and female stereotypes. [1]

- **File: "challenge_motifs_classifications.dta"** contains the MTurks classification of the 87 motifs tagged by terms related to challenges and competitions (according to ConceptNet).

- **File: "trick_deceive_cheat_motifs_classifications.dta"** contains the MTurks classification in terms of antisocial behavior of the 281 motifs tagged by terms related to trick, deceive, and cheat (according to ConceptNet).

4) Related terms of a concept for each motif. Related terms come from ConceptNet.[2] we look at the top-10k most common terms and drop NLTK stop words.

   a. 2,564 motifs **File: "Concepts_Tagged_Per_Motif.dta"**

5) Frequency of motifs mentioning at least one term related to (i) each of the top 10k concepts from ConceptNet,[3] (ii) entries in the Moral Foundations Dictionary, and (iii) entries in the LIWC.

   a. 958 groups in Berezkin **File: "Concept_Frequencies_Berezkin_groups dta"**

[1] The words used to identify female characters are woman, girl, female, wife, lady, mother, daughter, stepmother, grandmother, sister, niece, girlfriend, queen, and princess., For male characters, the words used are man, boy, male, husband, father, son, stepfather, grandfather, brother, nephew, boyfriend, king, and prince.

[2] To construct a ConceptNet-based list of related terms we retrieve the top-50 list for each seed word using the Web API version: api.conceptnet.io.

[3] Note that NLTK stop words are excluded and we omit keywords that have no words tagged in the motifs. So, instead of 10,000 concepts we end up having 9,005 concepts.

b. 1,245 groups in the EA **File: "Concept_Frequencies_EA_groups dta"**

c. 199 countries **File: "Concept_Frequencies_countries.dta"**

6) To help researchers and disseminate the ideas more widely we also make available the:

a) Population shares of Berezkin groups per country in 2000.

File: "Berezkin_to_Country.dta"

b) To arrive at the share of the country's population in 2000 of a Berezkin group, we matched the groups in Berezkin to those in the Atlas Narodov Mira. The following file contains the correspondence:

File: "Berezkin_to_Atlas_Narodov_Mira_Match.dta"

Of course, the matching is not perfect and one may come up with better matches. So, we have included a *notes* variable in the dataset where we describe the rationale for each match. This way researchers can improve upon a match.

Details on Matching: As we describe in the main text: "To arrive at a catalogue of folklore across countries, we first match the oral traditions listed in Berezkin to the ethnic groups in the 1964 Atlas Narodov Mira (ANM). Second, we use a group's fraction in the country's population in 2000 (combining the ANM polygons with information on cell-level population density from the Gridded Population of the World within country borders as of 2000) as a weight in our aggregation of motifs at the country level. This procedure is accurate for countries where the indigenous population as of 1500 constitutes a significant fraction of

the 1960 population (when the ANM dataset was built). For the countries, however, that ANM lists ethnicities that appeared after 1500, like the Brazilians, Australians, US Americans, etc., we use the population percentages from the countries of origin as reflected in the Putterman and Weil (2010) global migration matrix."

This country-of-origin adjustment affected a total of 28 countries mainly in the Americas. The post-1500 groups listed in the ANM are the: Panamans, Venezuelans, Uruguayans, US Americans, Salvadorians, Puerto Ricans, Peruvians, Nicaraguans, New Zealanders, Mexicans, Jamaicans, Hondurans, Haitians, Guatemalians, French-speaking population of Guiana, French Canadians, English Canadians, English-speaking population of Guiana, Ecuadorians, Dutch-speaking population of Guiana, Dominicans, Cubans, Costa Ricans, Columbians, Chileans, Brazilians, Bolivians, Australians, and Argentinians.

Caveat: To arrive at the country level population share of each group in the ANM as of 2000, we intersected the ANM shapefile with the country boundaries of 2000 depicted in the Digital Chart of the World. This projection was necessary as borders particularly in the regions of the ex-Soviet Union and the Balkans were redrawn between 1960 (when the ANM maps where drawn) and 2000. As a result of this intersection, tiny areas along country borders erroneously reflect groups that are not part of the current day countries. This is inconsequential since the weights assigned to these groups are infinitesimal.

Brown University, 64 Waterman Street, Providence, RI 02912. TEL: 401 863-3836.Email: smichalo@brown.edu

Section2

Replication Files for Tables and Figures that appear in the Manuscript **(folder: "Replication_Tables_Figures")**

First, make sure you install the following routines in Stata:
findit lassopack; findit rancktest; findit rancktest; findit acregpackcheck; findit geonear; findit geodist; findit binscatter

There are 5 do files and 4 data files.

1) The Tables and Figures that correspond to the country-level analysis can be regenerated by running the **"Country_Tables_Figures_FINAL.do"**

Figures: IVb, IVc, IVd, IIIb, III, VIIb; Tables: V (Panel A), VI (columns 1-8), VII (Panel A)

Appendix Figures: 10b; Appendix Tables: 6, 7 Panels A & B, 9a, 9b, 10

2) The Tables and Figures that focus on the groups in the Ethnographic Atlas can be regenerated by running the **"Ethnographic_Atlas_Tables_Figures_FINAL.do"**

Figures: Vc, VIIIb, VIIIc; Tables: II, III, IV, V (Panel C-columns 3 - 6), VII (Panel C);

Appendix Figures: 6a-6d, 7a-7c, 8a-8b, Appendix Tables: 2-Panel B, 4,

3) The Tables and Figures that focus on the groups in Berezkin's catalogue can be regenerated by running the **"Berezkin_groups_Tables_Figures_FINAL.do"**

Figures: II, IIIb, IIIc, Vb; Tables: I, V (Panel C – columns 1 & 2);

Appendix Figures: 1b, 4a-4e, Appendix Tables: 1, 2-Panel A, 3-Panels A, B, and C, 5.

4) The Tables and Figures that use 2nd generation immigrants from the European Social Surveys can be regenerated by running the

   "European_Social_Surveys_Tables_Figures_FINAL.do"

Figures: -; Tables: V (Panel B), VI (columns 9 & 10), VII (Panel B)

Appendix Figures: 10c; Appendix Tables: -

5) Appendix Table 8 can be reproduced by running the "**Appendix_Table_8.do**"