



Tabulated nonsense? Testing the validity of the Ethnographic Atlas

Duman Bahrami-Rad ^{a,*}, Anke Becker ^{b,1}, Joseph Henrich ^{a,1}



^a Department of Human Evolutionary Biology, Harvard University, United States of America

^b Harvard Business School, United States of America

ARTICLE INFO

Article history:

Received 17 March 2021

Received in revised form 21 April 2021

Accepted 23 April 2021

Available online 5 May 2021

JEL classification:

N01

N30

Z10

Z13

Keywords:

Ethnographic Atlas

Validation

Culture

ABSTRACT

The *Ethnographic Atlas* (Murdock, 1967), an anthropological database, is widely used across the social sciences. The *Atlas* is a quantified and discretely categorized collection of information gleaned from ethnographies covering more than 1200 pre-industrial societies. While being popular in many fields, it has been subject to skepticism within cultural anthropology. We assess the *Atlas's* validity by comparing it with representative data from descendants of the portrayed societies. We document positive associations between the historical measures collected by ethnographers and self-reported data from 790,000 individuals across 43 countries.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The *Ethnographic Atlas* (Murdock, 1967) is an anthropological database that contains information on more than 1200 pre-industrial societies. The data comes from ethnographies, i.e., rich and systematic portraits of diverse societies with information on various aspects of life, such as subsistence, political organization, or social norms. In essence, the *Atlas* is an effort to code, categorize and compile ethnographic information.

Across the social sciences, the *Atlas* has been widely used as a source of information about life in pre-industrial societies. More than twenty published articles in economics alone rely on it.² However, the *Atlas's* validity has never been systematically evaluated. This is surprising because its usefulness and data quality have been questioned in anthropology, the discipline from which it originated (Tobin (1990) provides an overview of criticism of the *Atlas* in cultural anthropology). While some, in the spirit of Boas (1896), denigrate the comparative nature of the database as 'tabulated nonsense' and 'grossly misleading' (Leach, 1964), others mention concerns about its historical validity (Jerven, 2011). Moreover, because the data for each society in the

Atlas is based on observations of at most a few ethnographers, who were typically male and European, researchers worry about the noisiness and biasedness of the data.

Perhaps the ideal test of the *Atlas's* validity would be representative data from the portrayed societies collected at the time when the ethnographers were making their observations. Unfortunately, such data does not exist. We therefore take the next best available data to assess its validity: data from descendants of the populations portrayed in the *Atlas*. This data comes from the *Standard Demographic and Health Surveys (DHS)*, representative household surveys that have been conducted recurrently in developing countries since 1984. Our sample of respondents from the *DHS* comprises 790,000 individuals from 317 ethnicities in 42 countries.

Our validation exercise is simple and straightforward. We identify all variables in the *DHS* that have an equivalent ethnicity-level counterpart in the *Atlas*. For these dimensions, we find positive associations between the historical information reported by ethnographers and the contemporary information reported by a large number of individuals. Importantly, the associations between historical ethnicity-level measures and contemporary self-reported data do not only hold for dimensions that would have been easy to observe for an ethnographer, such as how much a society relies on agriculture, or whether marriages are polygynous. Rather, they also hold for dimensions that are more concealed, such as how long couples abstain after birth, or whether people prefer sons. Moreover, the time lag between the historical and the contemporary data and the implied societal change suggests that

* Corresponding author.

E-mail addresses: duman_b_rad@fas.harvard.edu (D. Bahrami-Rad), a.becker@hbs.edu (A. Becker), henrich@fas.harvard.edu (J. Henrich).

¹ Anke Becker and Joseph Henrich thank CIFAR, Canada for generous funding.

² The supplementary material provides a list of scientific articles that use data from the *Atlas* published in the last two decades from diverse disciplines.

the associations we find can be interpreted as a lower bound of the true informational value of historical ethnographic accounts.

Our results document the validity and reliability of the *Atlas*. This is valuable not only for researchers using this database, but for anyone relying on historical or ethnographic data more generally.

2. Data

2.1. Data sources

The Ethnographic Atlas. The *Atlas* is a historical dataset (Murdock, 1967) with information on more than 1200 pre-industrial societies, covering various aspects of life, such as subsistence mode, kinship organization, religious beliefs, or political organization. This information is taken from ethnographies, i.e., systematic, extensive and detailed descriptions of a society by an ethnographer. For a small subset of societies, the so-called *Standard Cross-Cultural Sample* (SCCS) (Murdock and White, 1969) even more detailed information is available.

The Standard Demographic and Health Surveys. The *Standard Demographic and Health Surveys* (DHS) are nationally representative household surveys covering more than 90 countries worldwide. Respondents are women as the DHS aims to assess the life and well-being of women and children in developing countries. Until today, seven waves have been conducted, starting in 1984. The DHS country samples are quite large, with typically between 5000 up to 30,000 households being surveyed. Next to standard sociodemographic information, the DHS records information at the individual level about some dimensions that are included at the aggregate (ethnicity) level in the *Atlas*, such as whether a household is polygynous, how long a child was breastfed, or the geographic location of the household. We make use of this overlap to compare the historical information that was typically collected by a single ethnographer to the information that large samples of individuals report about their lives today.

2.2. Linking the DHS to the Atlas

We match pre-industrial societies in the *Atlas* to individuals in the DHS who are their descendants based on common language. Section 3 in the supplementary materials describes the methodology we used and introduces the Database of Global Cultural Evolution, which makes accessible the matching used here.

3. Benchmarking the *Atlas* with the DHS

For our analysis, we included all individuals from the DHS to whom we could assign their ancestral pre-industrial society. This resulted in a total sample of 791,086 individuals from 43 countries who are descendants of 317 different societies portrayed in the *Atlas*, interviewed between 1986 and 2016.

We systematically checked the DHS questionnaires for all measures that are equivalent to *Atlas* variables. This gave us a well-defined ex ante criterion on which variables to include in our analysis. We identified twelve variables in the DHS that have direct counterparts in the *Atlas*.

3.1. Empirical strategy

Our validation exercise is simple and straightforward: we regress the individual-level measure from the DHS on the respective ethnicity-level measure from the *Atlas* and include country-fixed effects whenever we have enough within-country variation. Our regression specification is

$$y_{i,j}^c = \alpha + \beta \times y_j^h + \text{CountryFE} + \epsilon_{i,j}$$

with $y_{i,j}^c$ an outcome for individual i from society j in the DHS, y_j^h the corresponding measure of society j from the *Atlas*, CountryFE a set of country-fixed effects, and $\epsilon_{i,j}$ the error term.³

3.2. Results

Twelve domains are equivalently represented in the DHS and the *Atlas*: (1) patrilocality, (2) matrilocality, (3) polygyny, (4) reliance on animal husbandry, (5) reliance on agriculture, (6) length of post-partum abstinence, (7) breastfeeding duration, (8) insistence on virginity, (9) a preference for sons, (10) prevalence of domestic violence, (11) age difference between husband and wife, and (12) geographical location.³

Throughout, we find positive associations between the ethnographic information from the *Atlas* and the self-reported individual-level data from the DHS (Table 1).

Columns (1) to (5) list the results for variables that capture different aspects of kinship organization and subsistence modes. Almost all associations are positive, statistically significant, and sizeable. For example, a one standard deviation increase in the prevalence of historical patrilocality is associated with a 0.8 percentage points increase in the likelihood that an individual lives patrilocally today. This amounts to about twelve percent of the unconditional probability of living patrilocally in this sample (0.7). We can only speculate about the lack of association for reliance on agriculture, which could be due to differences in pre-industrial and contemporary agriculture, or the fact that the DHS variable captures only one specific aspect of contemporary reliance on agriculture.

Columns (6) to (11) list results for variables that capture social norms, customs, or preferences. Again, the associations between the historical and contemporary measures are positive throughout, in most cases statistically significant, and often meaningful in terms of size. For example, a one standard deviation increase in the historical length of post-partum abstinence is associated with a twelve percentage points increase in how long respondents today abstain after childbirth. For the preference of female virginity before marriage the association between the two measures is very small. This can plausibly be attributed to the lack of variation in the contemporary sample: about 93% of respondents express this attitude. Again, we can only speculate about the lack of association between the historical age of an infant at the onset of weaning in an ethnic group and the average breastfeeding duration of its descendants. It could be that male ethnographers could not make informed guesses about this dimension, or that breastfeeding practices have undergone substantial change during the past century.

Finally, we show that geographical location of the centroid of an ethnic group as reported by ethnographers is related to where people actually live today. For each individual in the DHS for whom we have information on geographical location, we calculate the distance in kilometers to the centroid of the homeland of her ancestral society. Figure 1 in the supplementary material shows the distribution. The median distance is 168 kilometers and a non-negligible fraction of about 12 percent live as close as 50 kilometers to the centroid of their ancestral homeland.

4. Conclusion

Across a wide range of dimensions, we document that accounts of ethnographers about life in pre-industrial societies as captured in the *Atlas* are informative. In most dimensions, they are significantly associated with self-reports of descendants of the portrayed societies. These associations can be interpreted as a lower bound. Thus, the *Atlas* is a meaningful source of information about diverse human societies.

³ Table 2 in the supplementary material describes how these dimensions are measured in the *Atlas* and the DHS.

Table 1
The association between historical and contemporary measures.

	Dependent variable:										
	Patrilocal Residence [0/1]	Matrilocal Residence [0/1]	Polygynous Marriage [0/1]	Owns Animals [0/1]	Owns Land for Agric. [0/1]	Post-Partum Abstinence [Std.]	Breastfeeding Duration [Std.]	Virginity Important [0/1]	Pref. for Sons [Std.]	Experienced IP-Violence [0/1]	Age Diff. Husb.-Wife [Std.]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Patrilocality [Std.]	0.0079*** [0.0023, 0.014]										
Matrilocality [Std.]		0.0022** [0.00050, 0.0039]									
Polygyny [Std.]			0.021** [0.0035, 0.039]								
Dep. on Animal Husb. [Std.]				0.042*** [0.021, 0.063]							
Dep. on Agriculture [Std.]					0.0040 [-0.036, 0.044]						
Duration of PP-Taboo [Std.]						0.12*** [0.044, 0.20]					
Age: Onset of Weaning [Std.]							0.0033 [-0.087, 0.094]				
Insist. on Virginity [Std.]								0.0083** [0.00012, 0.016]			
Son Preference [Std.]									0.036** [0.0073, 0.065]		
Wife Beating Common [0/1]										0.044** [0.0083, 0.079]	
Age Diff. Husb.-Wife [Std.]											0.12*** [0.036, 0.21]
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations	497167	496160	422485	485540	493452	164335	81320	44507	69283	18361	36863
# of Clusters	310	310	272	240	236	85	40	24	19	16	14
Mean of Dep. Var.	0.0671	0.0480	0.2914	0.5820	0.6237	0 (Std. Var.)	0 (Std. Var.)	0.9250	0 (Std. Var.)	0.2735	0 (Std. Var.)

Notes. OLS estimates. Standard errors clustered at the society level. Columns are organized by number of clusters. Confidence intervals in parentheses. Columns 8 to 11: bootstrapped standard errors. Country fixed effects not included in columns 9 to 11 because of lack of within-country variation. Sample sizes vary: (i) the DHS does not elicit every domain in every survey, (ii) information on some domains is available only for a subset of societies in the *Atlas*. The historical variables in columns 9 to 11 are part of the SCCS (see Appendix B in the appendix for details.) Table 2 in Appendix A in the appendix shows the results with standard errors clustered at the society and the country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2

The association between historical and contemporary measures: Two-way clustered standard errors.

	Dependent variable:										
	Patrilocal Residence [0/1]	Matrilocal Residence [0/1]	Polygynous Marriage [0/1]	Owes Animals [0/1]	Owes Land for Agric. [0/1]	Post-Partum Abstinence [Std.]	Breastfeeding Duration [Std.]	Virginity Important [0/1]	Pref. for Sons [Std.]	Experienced IP-Violence [0/1]	Age Diff. Husb.-Wife [Std.]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Patrilocality [Std.]	0.0079*** [0.0027,0.013]										
Matrilocality [Std.]		0.0022** [0.00039,0.0040]									
Polygyny [Std.]			0.021** [0.0046,0.038]								
Dep. on Animal Husb. [Std.]				0.042*** [0.025,0.060]							
Dep. on Agriculture [Std.]					0.0040 [-0.034,0.042]						
Duration of PP-Taboo [Std.]						0.12*** [0.048,0.20]					
Age: Onset of Weaning [Std.]							0.0033 [-0.083,0.090]				
Insist. on Virginity [Std.]								0.0070 [-0.0046,0.019]			
Son Preference [Std.]									0.036** [0.0051,0.066]		
Wife Beating Common [0/1]										0.043** [0.0100,0.077]	
Age Diff. Husb.-Wife [Std.]											0.12*** [0.040,0.19]
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations	497167	496160	422485	485540	493452	164335	81320	44507	69283	18361	36863
# of Clusters	363	363	312	262	257	101	57	26	23	17	18

Notes. OLS estimates. Standard errors clustered at the society and the country level. Confidence intervals in parentheses. Columns 8 to 11: bootstrapped standard errors. Country fixed effects not included in columns 9 to 11 because of lack of within-country variation. Sample sizes vary: (i) the DHS does not elicit every domain in every survey, (ii) information on some domains is available only for a subset of societies in the *Atlas*. The historical variables in columns 9 to 11 are part of the SCCS (see [Appendix B](#) in the appendix for details.) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A. Additional tables

See Table 2.

Appendix B. The standard cross-cultural sample

The Standard Cross-Cultural Sample ([Murdock and White, 1969](#)) is a subset of 186 particularly well-documented societies in the Ethnographic Atlas. Here, we list the societies that are included in our analysis, separate for each outcome.

Son preference. Nama, Tsonga, Bemba, Luapula, Banen, Igbo, Akyem, Afikpo, Ashanti, Fante, Masa, Konso, Kafa, Amhara, Wolof, Djafun, Asben, Aulliminden, Kazakh, Khmer, Annamese, Jakaltek, Inca.

Wife beating. Tsonga, Bemba, Luapula, Ganda, Igbo, Afikpo, Yako, Bambara, Maasi, Konso, Esa, Somali, Amhara, Djafun, Burusho, Ifugao, Jakaltek, Aymara.

Age difference between husband and wife. Nama, Bemba, Luapula, Igbo, Akyem, Afikpo, Ashanti, Fante, Masa, Konso, Kafa, Amhara, Wolof, Djafun, Asben, Aulliminden, Kazakh, Khmer.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.econlet.2021.109880>.

References

- Boas, Franz, 1896. The limitations of the comparative method of anthropology. *Science* 4, 901–908.
- Jerven, Morten, 2011. A clash of disciplines? Economists and historians approaching the African past. *Econ. Hist. Dev. Reg.* 26, 111–124.
- Leach, Edmund, 1964. Reply to Raoul Naroll's on ethnic unit classification. *Curr. Anthropol.* 5, 283–312.
- Murdock, George P., 1967. *Ethnographic Atlas*. University of Pittsburgh Press.
- Murdock, George P., White, Douglas R., 1969. The standard cross-cultural sample. *Ethnology* 8, 329–369.
- Tobin, Joseph, 1990. The HRAF as radical text? *Cult. Anthropol.* 5, 473–487.