

## Codebook for historical war practices datasets 1-3

Variable: Population size

Dataset variable name: pop\_res1

Description: Population estimates for each society at the date range being coded, binned according to the number of places.

Codes: (0-99=1; 100-999=2; 1000-9999=3; 10000-99999=4; 100000-999999=5; 1000000-9999999=6; 10000000-99999999=7; 100000000-999999999=8)

Transformations for analyses: 1-8, standardized (mean centered at 0, sd = 1)

Datasets: 1, 2, 3

Variable: Agricultural intensity

Dataset variable name: agr\_ins

Description: Ethnographic Atlas variable 28, ranging from no agriculture to agriculture with widespread irrigation systems

Original codes: 1 (no agriculture), 2 (occasional/casual), 3 (shifting/seasonal), 4 (horticulture), 5 (permanent agriculture), 6 (permanent agriculture with irrigation)

Codes for analyses: 1-6

Datasets: 1, 3

Variable: Political expansion

Dataset variable name: ter\_res

Description: Whether warfare conducted by the society increased the territory or population under its political control. This should be roughly equivalent to the SCCS variable 909:

Subjugation of territory or people and the Seshat warfare variable of Annexation. It indicates deliberate attempts to expand political control but not decentralized expansion into land e.g. for hunting, grazing, or farming that involves fighting.

Original codes: inferred/yes, inferred/no, unclear

Codes for analyses: 0/1

Datasets: 2, 3

Variable: Degree of political centralization

Dataset variable name: pol\_cent

Description: How centralized is the cultural unit to which the practices coded for the most external level of war refer.

Original codes: 1. acephalous (absence of any political organization, even locally; equivalent to EA090); 2. autonomous communities (local political integration transcending kin groups, includes societies coded under EA090 autonomous local communities and peace groups); 3. loose/nominal/minimal states (a central government with some control or authority over relatively autonomous regional rulers, equivalent to EA090 minimal states and Seshat nominal, loose polities); 4. confederated/feudal states (the central government has more control over regional rulers but not complete authority); 5. unitary states (as defined by the Seshat variable- regional governors are appointed and removed by the central authorities, taxes are imposed by, and transmitted to the center)

Codes for analyses: 1-5

Datasets: 1, 2, 3

Variable: Hierarchical governing levels

Dataset variable name: ext\_hier

Description: Number of administrative/governing levels beyond the local community; roughly equivalent to Ethnographic Atlas variable 33 'Jurisdictional hierarchy beyond local community' generally used as a measure of political complexity in cross-cultural studies.

Original codes: Coded from 0 to 5 rather than the original 1 to 5 in EA033 to make code equal to the number of levels and to accommodate larger polities in the dataset. From description on D-place: 0 (=EA033 1, acephalous, no political authority beyond community, autonomous bands and villages); 1 (one level, petty chiefdoms); 2 (two levels, larger chiefdoms); 3 (three levels, states); 4 (four levels, large states); 5 (my addition, five or more levels, empires)

Codes for analyses: 0-5, standardized (mean centered at 0, sd = 1)

Datasets: 1, 2, 3

Variable: Community size

Dataset variable name: comm\_size

Description: The average size of local communities, coded based on EA variable 031 'Mean size of local communities'

Original codes: 0 (<50), 1 (50-59), 2 (100-199), 3 (200-399), 4 (400-1000), 5 (1000-5000), 6 (5000-50000), 7 (50000+)

Codes for analyses: 0-7, standardized (mean centered at 0, sd = 1)

Datasets: 1

Variable: Political complexity

Dataset variable name: PC1

Description: Overall measure of political complexity created from community size (Chapter 1 only), population size, hierarchical levels, and centralization variables.

Original codes: NA

Codes for analyses: first component from PCA of above variables, standardized (mean centered at 0, sd = 1)

Datasets: 1, 2, 3

Variable: Presence of formal military organization

Dataset variable name: mil\_org

Description: Whether there is a formal standing army or military organization, or whether warriors are mobilized through social relations such as friends or kin. Coded according to the SCCS variable 894: Form of military mobilization (originally coded in Otterbein 1970).

Original codes: 0 (informal, social relations, mass mobilization), 1 (age-grades or armies with professional training)

Codes for analyses: 0/1

Datasets: 1, 2, 3

Variable: Self-sacrificial behavior in war

Dataset variable name: sac\_scale

Description: Is there any evidence that fighters are commonly expected to risk or sacrifice their lives for their companions or the group as a whole?

Original codes: 1 (no indication that risking one's life or sacrifice on behalf of the group is devalued or discouraged), 2 (indication that risking one's life is valued but unclear about degree to which sacrifice on behalf of others is valued), 3 (indication that loyalty to and sacrifice for the group is valued but no explicit mention or proof of sacrificial acts/deaths), 4 (explicit description of combat deaths on behalf of others in the group and that this is socially valued).

Codes for analyses: 1, 2, 3, 4

Datasets: 1

Variable: Institutionalized trophy-taking

Dataset variable name: trophies

Description: Whether body parts are taken from enemies in an institutionalized or widespread manner, as a primary or expected outcome of war.

Original codes: yes (described in sources), no (described in sources as taboo or explicitly not practiced), inferred no (no evidence in context of otherwise detailed descriptions of combat so can be inferred absent)

Codes for analyses: 0/1

Datasets: 3

Variable: Enemies killed in external war

Dataset variable name: kill\_ext

Description: Whether enemy individuals in the following age/gender categories tend to be killed during war: male or female infants/toddlers; children; younger adults; older adults

Original codes: 1-8. 1 point is added for each age/gender category, with the minimum being 0 (none) and the maximum being 8 (everyone). For instance, the Nuer would be scored a 5 (male and female infants/toddlers + male young adults + male and female older adults). Ranges are used for some societies when there is ambiguity or overlap across categories.

Codes for analyses: 1-8

Datasets: 2

Variable: Enemies killed measurement error

Dataset variable name: kill\_se

Description: The difference between the mean and maximum/minimum of range of enemies killed for each society.

Original codes: NA

Codes for analyses: 0-3

Datasets: 2

Variable: date

Description: The time period to which the other variables coded refer to, based on the date ranges covered by the sources referenced. For historical societies this generally is the start and end dates of a particular political organization, and may therefore be approximate. For ethnographically studied societies it is generally taken from the coverage dates given in eHRAF for the ethnographies used to code the other variables.

Codes: Year range

Variable: pop\_source

Description: Sources for population estimates. 'eHRAF' indicates the society's cultural summary page; 'EA202' indicates variable 202 population size in the Ethnographic Atlas; 'Seshat' indicates the polity population variable on the society's Seshat page; other sources cited specifically