

# Standard Cross-Cultural Sample<sup>1</sup>: on-line

George P. Murdock and  
Douglas R. White  
University of Pittsburgh

This paper presents the first research results of the Cross-Cultural Cumulative Coding Center (CCCCC), a unit established at the University of Pittsburgh in May, 1968, with support from the National Science Foundation. It offers to scholars a representative sample of the world's known and well described cultures, 186 in number, each "pinpointed" to the smallest identifiable subgroup of the society in question at a specific point in time.

## PINPOINTING

The basic assumption underlying cross-cultural research is that the elements of any culture tend over time to become functionally integrated or reciprocally adjusted to one another. As new elements are invented or borrowed they are gradually fitted into the pre-existing cultural matrix, and the latter is modified to accommodate them. If such adaptations were instantaneous, cultures would at all times exhibit perfect integration, and, as British social anthropologists following Radcliffe-Brown have often rashly assumed, functional relationships among the elements of a culture or social system would readily become manifest through observation and analysis. If such were the case, valid scientific generalizations could be reached by the intensive study of individual cultures, and cross-cultural research would be unnecessary (a common corollary assumption among the same group of scholars).

Unfortunately, the integrative process is much more complex. New and introduced elements of culture require time—often the passage of generations—before they become adapted to the cultural matrix and fit to them. Ogburn (1922: 200-280) coined the apt term "cultural lag" for this necessary interval. Because of the delay in adjustment, cultures never reveal perfect integration but only what Sumner (1906: 5-6) called a "strain toward consistency," in which innumerable adaptive processes are continually being in-

---

<sup>1</sup> Reprinting of the 1969 sections of the article with this title (Ethnology 8:329-369) by permission of the journal editors. Page breaks are identified as per original article. Footnotes were added in 2002 by DRW to provide additional information, and links to the on-line dataset at <http://eclectic.ss.uci.edu/~drwhite/worldcul/Sccs34.htm> provide the list of societies in Table 1 and data on these societies Appendix A ([go there](#)).

errupted by innovations and other events before they have fully run their course. Thus any culture at any time exhibits relationships among its constituent elements which are in part completely integrated functionally and in part still unadjusted or only imperfectly adjusted to each other. It is impossible for ethnographic analysis to identify these except by conjecture—and the conjectures of anthropologists who investigate functional relationships are not notably superior to those of anthropologists who attempt to reconstruct culture history.

329

### 330 ETHNOLOGY

When faced with problems of this sort, scientists resort to statistics. Distrustful of ad hoc interpretations of single instances, they examine a large and representative number of cases to determine whether the postulated relationships among relevant variables are quantitatively substantiated. This is precisely the rationale of cross-cultural research. Since some of the relationships among cultural elements are surely functional whereas others are not, statistics offers the only dependable technique for segregating them and thus arriving at scientifically valid generalizations.

By definition, cultural elements can be considered functionally interrelated only if they occur together at the same time among the same culture-bearing group. Elements occurring in the same society at different time periods, or in culturally variant subgroups, cannot be assumed to be functionally related, even though in some cases they may still reflect an earlier functional congruence. Methodologically, it is just as crucial to establish the actual concurrence of elements as to employ appropriate sampling and other statistical techniques to the analysis of their association. Unfortunately, previous cross-cultural research has commonly been as defective with respect to the former as to the latter requirement. The Human Relations Area Files assembles materials under the name of a society regardless of the date of the ethnographic observations and of whether the name embraces subgroups with marked cultural differences, and few researchers have sought a sharper focus, even though HRAF normally supplies the necessary information. All too frequently, therefore, assumed associations or their negation are invalidated by differences of time or place between the elements compared.

The authors have made a strenuous effort to correct this defect. They have "pinpointed" every society in the standard sample to a specific date and a

specific locality, typically the local community where the principal authority conducted his most intensive field research, and they have eliminated from the sample all societies for which the sources do not permit reasonably accurate pinpointing. It is often possible, of course, for coders to make fairly reliable inferences by extrapolation from data on neighboring local groups with closely similar subcultures or from observations at somewhat earlier or later dates, but they should be instructed to use great caution in so doing and to specify their reasons. Focusing attention on a specific pinpointed date and locality should, it is believed, substantially enhance the accuracy of cross-cultural research.

Appendix A summarizes the data on the pinpointing sheets prepared by the Cross-Cultural Cumulative Coding Center for the guidance of its coders.<sup>2</sup> The appendix specifies for each society of the standard sample its geographical focus with the name and coordinates of the pinpointed subgroup, its temporal focus with a specific date, and some indication of the reasons for the selection of both. It also names, for the guidance of researchers, the authority or authorities adjudged most dependable or useful, indicates whether or not HRAF has a file on the society (giving its identifying number) and whether such a file is of excellent, good, or marginal quality (i.e., the extent to which supplementation by library research may be necessary), and presents in abbreviated form the information on linguistic

#### STANDARD CROSS-CULTURAL SAMPLE 331

affiliation, type of subsistence economy, level of political integration, and prevailing rule of descent used below in estimating the relative strength of historical and integrative factors in producing cultural similarities. Bibliographical references to the principal sources on most of the societies may be found in the Ethnographic Atlas of this journal by means of the identifying numbers of the societies as given in the Appendix. Supplementary information will appear in future installments of the Atlas.<sup>3</sup>

---

<sup>2</sup> Appendix A is omitted but corresponds to coded data in *World Cultures* files STDS34.SAV (STDS34.DAT needs to be updated) with variables 833-844 explained in <http://eclectic.ss.uci.edu/~drwhite/worldcul/ScCS34.htm> ([go there](#)) and on-line codebook, <http://eclectic.ss.uci.edu/~drwhite/worldcul/SCCS1.htm> ([go there](#)). An undergraduate class using these materials on-line is found at <http://eclectic.ss.uci.edu/~drwhite/courses/WCC03.html> ([go there](#)).

<sup>3</sup> The Ethnographic Atlas is now on-line through *World Cultures* and the Center for Anthropological Computing at the University of Kent at ([go there](#)).

## SAMPLING

Another major but vexing problem of cross-cultural research is that of sampling, i.e., the selection for comparison of a number of societies that will adequately represent the entire range of known cultural variation and at the same time eliminate as far as possible the number of cases where similarities are presumably due to the historical influences of diffusion or common derivation. Five methods of coping with "Galton's problem," as this is called, have been advanced by Naroll (1961, 1964; Naroll and D'Andrade 1963), several of which the present authors have found useful.

Since his first major work utilizing a large cross-cultural sample (Murdock 1949), the senior author has proposed a series of improved and expanded world samples (Murdock 1957, 1963, 1968) and has discussed the problem of sampling from a theoretical point of view (Murdock 1966, 1968). The standard cross-cultural sample presented herewith represents the culmination of these efforts. It is designed specifically to correct a serious short-coming of previous cross-cultural research. In general, each scholar has worked with a sample of his own choosing. Most of these have been small and selected in a relatively haphazard fashion with only casual reference to sampling principles.<sup>4</sup> The selections have varied with each researcher, and the overlap between them has been small, so that it has rarely been possible to intercorrelate the findings of different studies. The progress of cross-cultural research has consequently been slow, and merely additive rather than multiplicative. What is needed to correct this situation is a large world sample, constructed with strict regard to ethnographic distributions and sophisticated sampling procedures, which can be used in different studies, so that the results of each can be intercorrelated with one another and the progress of comparative research thus lifted from an arithmetic to a geometric rate.

The establishment of such a sample has depended on three arduous but necessary preliminary research activities: (1) the analysis of more than 1,250 societies (in the Ethnographic Atlas of this journal), a very high proportion of all those whose cultures have been adequately described, to identify those with the fullest ethnographic coverage and to make certain that no major cultural variant has been overlooked; (2) the classification of all the cultures assessed into "clusters" (Murdock 1967), i.e., groups of contiguous

---

<sup>4</sup> Ref to DRW 1965 on sample construction: and the HRAF Probability Sample (Naroll 1967; Lagacé 1979), which is a subset of the HRAF Collection of Ethnography ([go there](#)).

societies with cultures so similar, owing either to diffusion or to recent common origin, that no world sample should include more than one of them; (3) the grouping of clusters, usually but not always adjacent, into "sampling provinces" (Murdock 1968) where linguistic and cultural evidence reveals

### 332 ETHNOLOGY

similarities of a lesser order but still sufficient to raise the presumption of historical connections in violation of Galton's objections.

In general, one society was selected from each of the 200 world sampling provinces distinguished by Murdock (1968). However, for two of these provinces—Ancient Egypt (50) and Tasmania (110)—pinpointing to a particular locality and date proved impossible, and they are therefore unrepresented in the standard sample. Two other provinces (56 and 159) have been split in half and a representative selected from each. In fourteen other instances (Provinces 16, 26, 29, 33, 47, 71, 74, 87, 125, 147, 148, 157, 175, and 181) no representative was selected because the province was adjudged to resemble another too closely to warrant representation of both. In view of these changes and of the numerical reordering of the provinces for reasons to be presented later, the revised provinces from which the 186 sample societies are drawn will be designated hereinafter as "distinctive world areas" (or simply "areas") to distinguish them from the original "world sampling provinces."

The 186 societies included in the standard sample are distributed relatively equally among the six major regions of the world, as follows:

A (Sub-Saharan Africa)	28
C (Circum-Mediterranean)	28
E (East Eurasia)	34
I (Insular Pacific)	31
N (North America)	33
S (South and Central America)	32

If Africa appears slightly underrepresented, the reader should note that the sample also includes two Negro societies in South America as well as several on the Sudan fringe of the Circum-Mediterranean.

The selection of the particular society to represent an area was based in most cases on the adjudged superiority of its ethnographic coverage. Sometimes, however, the overriding criterion was its distinctiveness in world perspective as regards either language (as in Areas 50 and 64), economy (as in 9 and 93), political organization (as in 31 and 54), or descent (as in 4 and 101). Oftentimes, too, the choice was determined by the availability of information in the Human Relations Area Files (as in Areas 55, 80, 81, 89, and 140). The standard sample thus constructed will naturally be modified in detail as superior ethnographies become available and as historical relationships within and between the sampling provinces are clarified by future research. For the time being, however, it represents the best selection of which the authors are capable. Comparison of our standard sample with previous samples of comparable or larger size (see Table 1) reveals, it is believed, its definite superiority in range, exhaustiveness, and relative independence of cases.

The sample of 250 societies used in *Social Structure* (Murdock 1949), when judged by current standards, was so obviously defective in several important respects that it can be excused only as a pioneer effort. The societies were not pinpointed in time or space. Two of the great world regions—the Circum-Mediterranean and South America—were seriously under-

#### STANDARD CROSS-CULTURAL SAMPLE 333

TABLE 1 (STANDARD SAMPLE with Comparisons)

---



---

Data omitted on pp. 333-335 are available on the web  
at <http://eclectic.ss.uci.edu/~drwhite/worldcul/Scs34.htm> ([go there](#)).

#### 336 ETHNOLOGY

TABLE 1 (*continued*)

Data omitted at the top of p. 336 are available on the web

represented, as were complex societies in general. No examples were included from the past civilizations that are adequately described in contemporary documents. For the present sample the authors made a special

search for such cases and were able to include three—the Babylonians at the end of Hammurabi's reign (Area 45), the Hebrews at the time of the promulgation of the Deuteronomic code (Area 44), and the Romans of the early imperial period (Area 49), for whom the Athenians of the Periclean Age might well have been substituted. Several other possibilities were rejected as inadequately described or incapable of exact pinpointing as to date or locality, notably the Carthaginians, Egyptians (of either the Middle Kingdom or the New Empire), Indo-Aryans, Persians, Scythians, and Sumerians, although an early dynastic period in China could probably have been used if the authors had been able to control the Chinese language and literature.

The numerous gaps and duplications of the SS sample constitute an even more serious fault. Of the 186 areas represented by one case each in the present sample, 75 (or 41 per cent) had no representatives in the earlier sample,

#### STANDARD CROSS-CULTURAL SAMPLE 337

whereas 34 (or 18 per cent) were heavily overrepresented, i.e., by three or more cases each. Fortunately it is possible to profit by one's own past errors.

The sample of 400 societies used by Textor (1967), derived from the Ethnographic Atlas (Murdock 1963), reveals a marked improvement. Of the 186 areas of our standard sample, only eleven (or 6 per cent) are completely unrepresented and 28 (or 15 per cent) are seriously overrepresented. Textor's correlations can therefore be accepted with only modest reservations. The chief lesson to be learned from Textor's work, however, is that samples of such large size are unnecessary. Our own research indicates that a carefully drawn sample of around 200 cases essentially exhausts the universe of known and adequately described culture types.

The files of good (or at least fair) quality produced to date by the Human Relations Area Files constitute a third large world sample—and one deserving special analysis because of their enormous potential utility in cross-cultural research. From an examination of the set at the University of Pittsburgh the authors have estimated that the files on 220 societies are usable, though in some cases only marginally so. They are analyzed in Table 2 with regard to their distribution by major regions and distinctive world areas, by their use and availability for the societies of our standard sample, and by the relative quality of those available.

TABLE 2. DISTRIBUTION OF HRAF FILES BY REGIONS, AREAS, QUALITY, AND USE

Major Region	Number of HRAF Files	Areas Represented By				Sample Societies		
		No Files	One File	Two Files	3-6 Files	Using HRAF Files Ade-quate	Mar-ginal	Societies Not Using Available HRAF Files
A	30	10	8	8	2	16	1	1
C	29	12	9	4	3	13	2	1
E	47	4	19	7	4	17	8	5
I	32	7	16	7	1	14	5	5
N	46	8	12	6	7	19	2	4
S	36	5	18	9	0	22	1	4
Total	220	46	82	41	17	101	19	20

In the adequacy of its distribution the HRAF sample stands about mid-way between the SS and Textor samples. It lacks any representative for 46 (25 per cent) of our 186 distinctive areas, ranging from a maximum of 12 (43 per cent) of those of the Circum-Mediterranean to a minimum of 4 in East Eurasia.<sup>5</sup> Of the societies in our standard sample, 120 (or 64 per cent) are represented by HRAF files; 19 of these, however, are adjudged only marginally useful for the pinpointed subgroup, i.e., requiring supplementation by library research. In 20 instances in which a HRAF file exists for a particular area, another society, not represented in HRAF, was selected as more suitable. Hence researchers may, without departing from strict sampling principles, substitute any of the following HRAF files for the alternative selections in the standard sample: Yoruba for Fon in Area 18, Maltese for

### 338 ETHNOLOGY

Romans in 49, Kol for Santal in 62, Khasi for Garo in 69, Kachin for Lakher in 70, Macassarese for Toradja in 87, Wogeo for Kwoma in 95, Kurtatchi (Buka) for Siuai in 99, Malekula or Santa Cruz for Pentecost in 101, Lau

<sup>5</sup> Since 1968, HRAF has made little effort to expand their ethnographic files to include the cases lacking from the standard sample. (go there)



Fijians for Mbau Fijians in 102, Okinawans for Japanese in 117, Yakut for Yukaghir in 120, Tlingit for Haida in 131, Mandan for Hidatsa in 141, Iroquois for Huron in 144, Tewa for Chiricahua in 148, Yucatec Maya for Quiche in 155, Tucuna for Cubeo in 167, Bacairi or Bororo for Trumai in 175, Caraja for Shavante in 179. They must, however, resort to library research for any society in the 46 areas not represented in HRAF; failure to do so would result in sampling distortion and consequent reduction in the reliability of findings.

Cross-cultural researchers, in short, though they will find HRAF of invaluable assistance, cannot depend exclusively on this resource if they wish to adhere to sophisticated sampling procedures. Little improvement in this respect can be envisaged for the future. The so-called "blue ribbon" enrichment sample of fifteen societies on which the HRAF staff is currently working, for example, will fill only three of the 46 areas now unrepresented: 27 (Kanuri), 50 (Bahia Brazilians), and 181 (Guarani). Four other cases are improvements of already existing files: Somali of Area 36, Serbs of 48, Khasi of 69, and Siamese of 76. The remaining eight will merely produce new files on areas already represented, adding the Dogon to the Mossi and Tallensi in Area 33, the Santal to the Kol in 62, the Garo to the Khasi in 69, the Sinhalese to the Vedda in 80, the Toradja to the Macassarese in 97, the Taiwan Chinese to four existing Chinese files in 114, the Tzeltal to the Yucatec Maya in 155, and the Tucano to the Tucuna in 167.

Most cross-cultural samples reveal a definite bias in favor of sources in the English language. In compiling our standard sample the authors have specifically sought to discount this bias and to select the best described societies irrespective of the nationality or native language of the ethnographers. To be sure, speakers of English (Americans, Australians, British, Canadians, New Zealanders, and South Africans) have made a preponderant contribution to world ethnography, and it is not surprising that they include the principal authorities on 62 per cent of the societies in the standard sample. However, the contributions from other nationals have been far from negligible. We estimate the number of societies in our sample for which the principal authority is (or was) a native speaker of a language other than English, regardless of the language in which his description may have been written or published, as follows:

4 Arabic or other Semitic	2 Italian or Latin
2 Bantu or other African	2 Japanese

3 Chinese  
4 Dutch  
16 French  
16 German

6 North Germanic or Scandinavian  
9 Russian or other Slavic  
5 Spanish  
1 Turkish

Excluded from the above count, of course, are authors who emigrated in childhood or youth from one language area to another.

In the effort to achieve diversity and relative historical independence

#### STANDARD CROSS-CULTURAL SAMPLE 339

among the societies of the standard sample, we selected a single society from each of 51 independent linguistic families (not necessarily independent phyla) and also one as the sole representative of each of 34 linguistic subfamilies where other subfamilies of the same family were represented. Since the speakers of distinct subfamilies have normally been separated for well over 2,000 years, the cultures of these 85 societies, constituting 43 per cent of the total sample, may be considered exempt or nearly so from historical influences stemming from common origins. The remaining 101 societies, not thus exempt, have been chosen from linguistic subfamilies having other representatives in the sample, as follows:

2 from the Berber, 2 from the Chadic, 4 from the Cushitic, and 5 from the Semitic subfamilies of Afro-Asiatic or Hamito-Semitic;  
4 from Algonkian;  
2 from the Turkic subfamily of Altaic;  
3 from the Northern subfamily of Athapaskan;  
2 from Australian;  
2 from Cariban;  
2 from the Eastern subfamily of Chari-Nile or Macro-Sudanic;  
2 from Chibchan (as well as a third from a different subfamily);  
2 from Dravidian;  
2 from Ge (as well as a third from a distinct subfamily);  
3 from the Indic, 2 from the Iranian, and 3 from the Italic or Romance subfamilies of Indo-European;  
2 from the Southern subfamily of Khoisan;  
4 from the Carolinian, 11 from the Hesperonesian, 6 from the Melanesian, and 4 from the Polynesian subfamilies of Malayo-Polynesian;  
2 from Nahuatlan or Mexicano;

2 from the Atlantic, 13 from the Bantoid or Central, 3 from the Kwa, and 2 from the Mande subfamilies of Niger-Congo;  
2 from Salishan;  
2 from Shoshonean;  
2 from Siouan;  
4 from Tupi-Guarani.

Such linguistic duplication is impossible to avoid in Africa, the Circum-Mediterranean, and the Insular Pacific because of the preponderance of very large families and subfamilies in these regions, and especial attention was therefore paid to cultural differentiation in selecting samples from these linguistic groups.

The distribution of other cultural features in the sample probably approximates that in the world as a whole. The major types of subsistence economy in the sample societies are classified as follows:

- 141 food-producing economies: 126 agricultural and 15 pastoral. In Appendix A, the agricultural economies are further differentiated into advanced agriculture (56), horticulture (19), and simple or swidden agriculture (51).
- 45 food-collecting economies: 13 primarily gathering, 14 primarily hunting (including 4 societies of equestrian hunters), and 18 primarily fishing (including one which depends for food more on trade than on direct subsistence activities). The great majority of the food-collecting societies are located in North and South America.

The complexity of political organization (see Table 3) is greatest in the circum-Mediterranean, intermediate in Africa and East Eurasia, and rel-

TABLE 3. RELATIVE POLITICAL COMPLEXITY BY MAJOR WORLD REGIONS

Major Region	Stateless Societies	Societies With Minimal States	Societies With Small States	Societies With Large States	Total
A	8	9	6	5	28
C	1	7	5	14	28
E	11	10	2	11	34
I	13	11	6	1	31
N	23	7	2	1	32
S	23	5	2	2	32
Total	79	50	23	34	186

atively slight in the Insular Pacific and the Americas. The several societies which form an integral part of a large state dominated by another society are classed as having the level of political integration of their rulers (i.e., with three or more levels above that of the local community).

The major types of social organization, as represented by the prevailing rule of descent, reveal a fairly even distribution among the regions of the earth (see Table 4), despite some preponderance of patrilineal forms in the Old World and of bilateral forms in the New World. In Table 4, quasi-patrilineal descent (three cases) is grouped with bilaterality, and double descent is classified as matrilineal (one case) or patrilineal (six cases) depending upon the prevailing rule of marital residence.

In selecting the dates for which the societies of the sample are pinpointed, the authors have in general chosen the earliest period for which satisfactory ethnographic data are available or can be reconstructed, though we have sometimes used a later date for which the descriptive materials are appreciably richer. The reason, of course, is to avoid insofar as possible the acculturative effects of contacts with Europeans, which in recent centuries have exerted a convergent influence on all the cultures of the world. To further offset this influence, we have excluded from our sample the great colonizing and imperialistic societies of Europe—the Belgians, Dutch, English, French, Germans, Italians, Portuguese, and Spaniards—since their cultures are assumed to be reflected to some degree in those of the peoples they have governed or missionized. Table 5 shows the range of the pinpointed dates for the societies of the sample, classified by major regions as

**TABLE 4. RULES OF DESCENT BY MAJOR REGIONS**

Region	Ambilineal	Bilateral	Matrilineal	Patrilineal	Total
A	2	3	4	19	28
C	0	9	2	17	28
E	0	9	5	20	34
I	6	7	7	11	31
N	2	20	8	3	33
S	0	21	5	6	32
Total	10	69	31	76	186

**STANDARD CROSS-CULTURAL SAMPLE 341****TABLE 5. RANGE OF PINPOINTED DATES FOR THE SAMPLE SOCIETIES**

TimePeriod	A	C	E	I	N	S	Total
1750to1B.C.	0	2	0	0	0	0	2
A.D.1to1500	0	1	1	0	0	0	7
A.D.1501to1600	0	0	0	0	1	2	3
A.D.1601to1700	0	0	0	0	2	1	3
A.D.1701to1800	0	0	0	1	3	1	5
A.D.1801to1850	0	1	0	4	4	2	11
A.D.1851to1900	9	7	11	4	16	5	52
A.D.1901to1950	19	13	17	16	7	19	91
A.D.1951to1965	0	4	5	6	0	2	17
Total	28	28	34	31	33	32	186

well as time periods. The earliest dates, of course, are for the Old World whereas the New World shows a preponderance of the intermediate dates. The heaviest concentration naturally falls in the century from 1851 to 1950, the heyday of professional anthropology.

**MAPS**

Six maps, each constructed to a similar scale on equal-area projections, locate the pinpointed focus of 184 of the sample societies; only the Marquesans and Samoans of Polynesia could not be accommodated. The authors acknowledge with gratitude the assistance rendered by Professor Hibberd V. B. Kline, Jr., and Mr. Howard N. Ziegler of the Department of

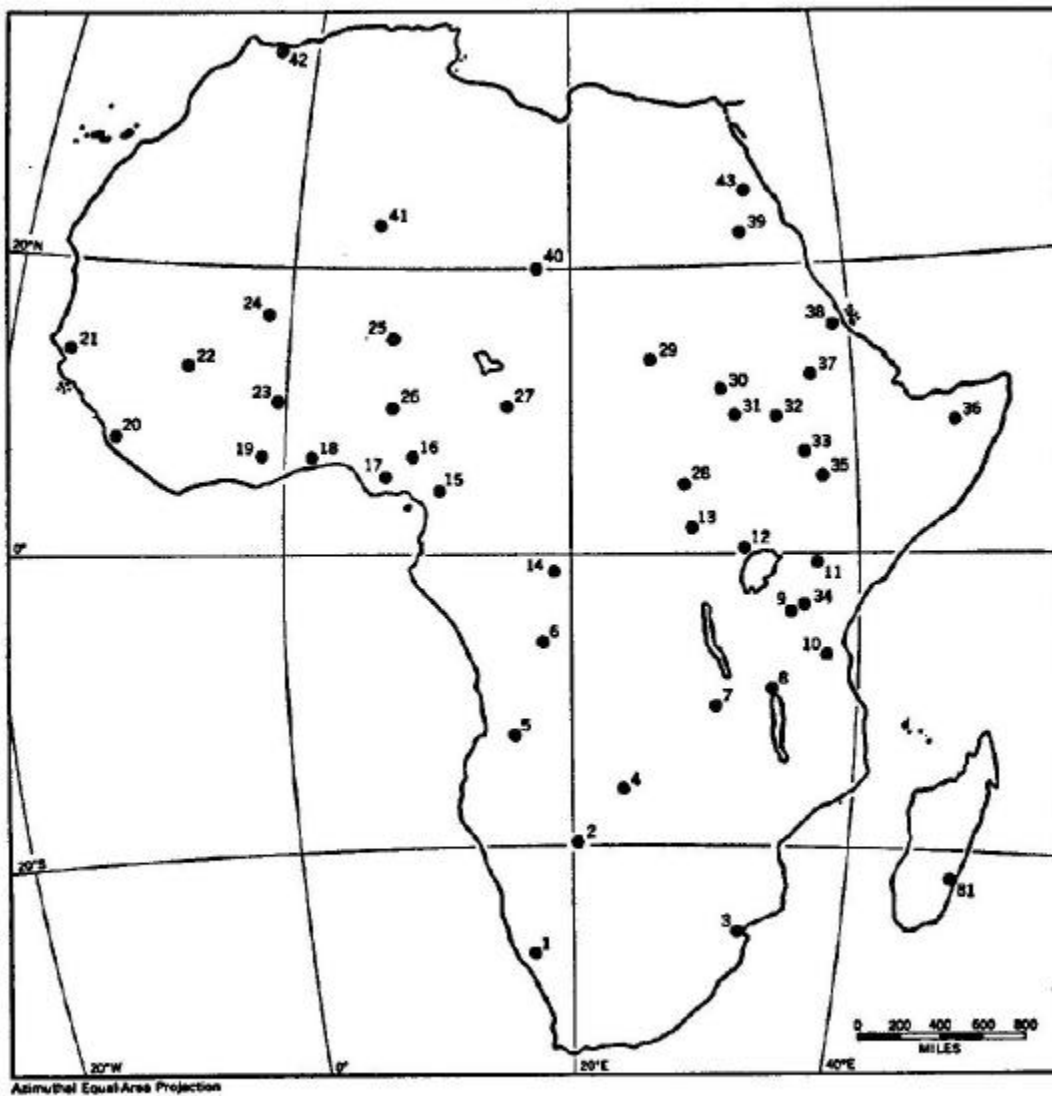
Geography, University of Pittsburgh, in preparing these maps. They give a graphic picture of the geographic distribution of the sample societies. The areas where such societies are sparsely represented are in most cases either those which are largely uninhabited (like much of the Sahara Desert and of Arctic Canada), those whose indigenous cultures mainly disappeared before they were recorded (like the eastern seaboard of Brazil and the United States), or those where essentially similar cultures cover extensive territories (as in China and aboriginal Australia). In general, only inhabited islands are indicated (outlined or suggested by dots), others being completely omitted, and in no case are areas shown in one map duplicated in any other. The maps correspond only roughly to our six major world regions (A, C, E, I, N, and S), the discrepancies being noted beneath each map.

The numerical order in which the 186 societies (and areas) of the sample have been arranged (see Table 1 and Appendix A) can be readily followed on the maps by observing their numbers sequentially. The ordering, though it may appear arbitrary at first glance, was designed to place each area, insofar as possible, between the two others to which it is geographically most contiguous and culturally most similar. The alignment thus zigzags across the maps, guided by the restraints imposed by major geographic, linguistic, and ethnic boundaries, and crosses from one region to another where they are most contiguous.

(societal names have been added on the pages of the maps)

1 Nama Hottentot	1860	16 Tiv	1920	31 Shilluk	1910
2 Kung Bushmen	1950	17 Ibo	1935	32 Mao	1939
3 Thonga	1865	18 Fon	1890	33 Kaffa (Kafa)	1905
4 Lozi	1900	19 Ashanti	1895	34 Masai	1900
5 Mbundu	1890	20 Mende	1945	35 Konso	1935
6 Suku	1920	21 Wolof	1950	36 Somali	1900
7 Bemba	1897	22 Bambara	1902	37 Amhara	1953
8 Nyakyusa	1934	23 Tallensi	1934	38 Bogo	1855
9 Hadza	1930	24 Songhai	1940	39 Kenuzi Nubians	1900
10 Luguru	1925	25 Pastoral Fulani	1951	40 Teda	1950
11 Kikuyu	1920	26 Hausa	1900	41 Tuareg	1900
12 Ganda	1875	27 Massa (Masa)	1910	42 Riffians	1926
13 Mbuti	1950	28 Azande	1905	43 Egyptians	1950
14 Nkundo Mongo	1930	29 Fur (Darfur)	1880		
15 Banen	1935	30 Otoro Nuba	1930	81 Tanala	1925

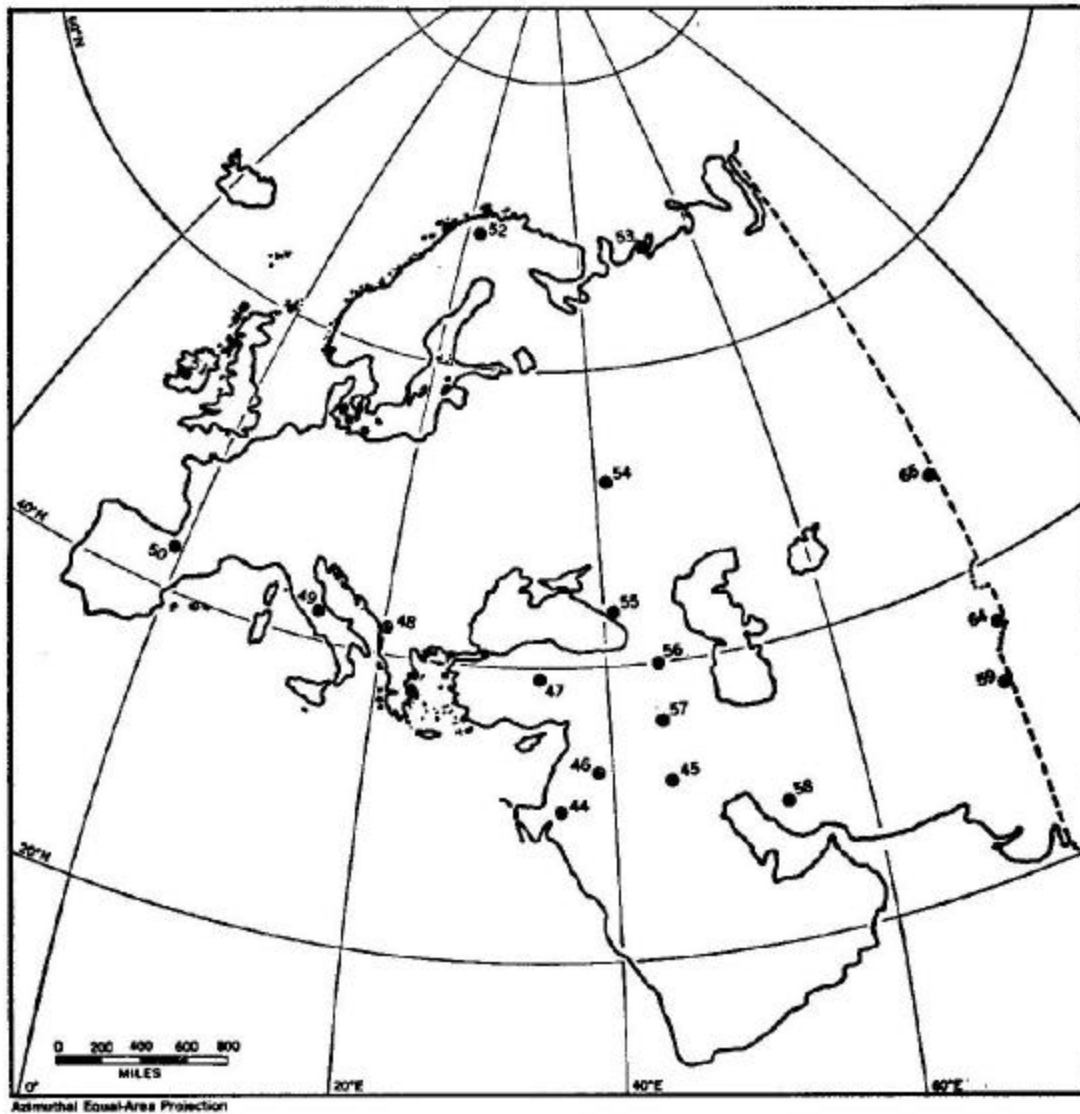
MAP I: AFRICA



This map locates all 28 of the sample societies in Sub-Saharan Africa (A), fifteen of those in the Circum-Mediterranean region (C), and one in Madagascar, which is included in East Eurasia (E).

44 Hebrews	- 621	53 Yurak Samoyed	1894
45 Babylonians	-1750	54 Russians	1955
46 Rwala Bedouin	1913	55 Abkhaz	1880
47 Turks	1950	56 Armenians	1843
48 Gheg Albanians	1910	57 Kurd	1951
49 Romans	110	58 Basseri	1958
50 Basques	1934	59 Punjabi (West)	1950
51 Irish	1932	64 Burusho	1934
52 Lapps	1950	65 Kazak	1885

MAP 2: WEST EURASIA

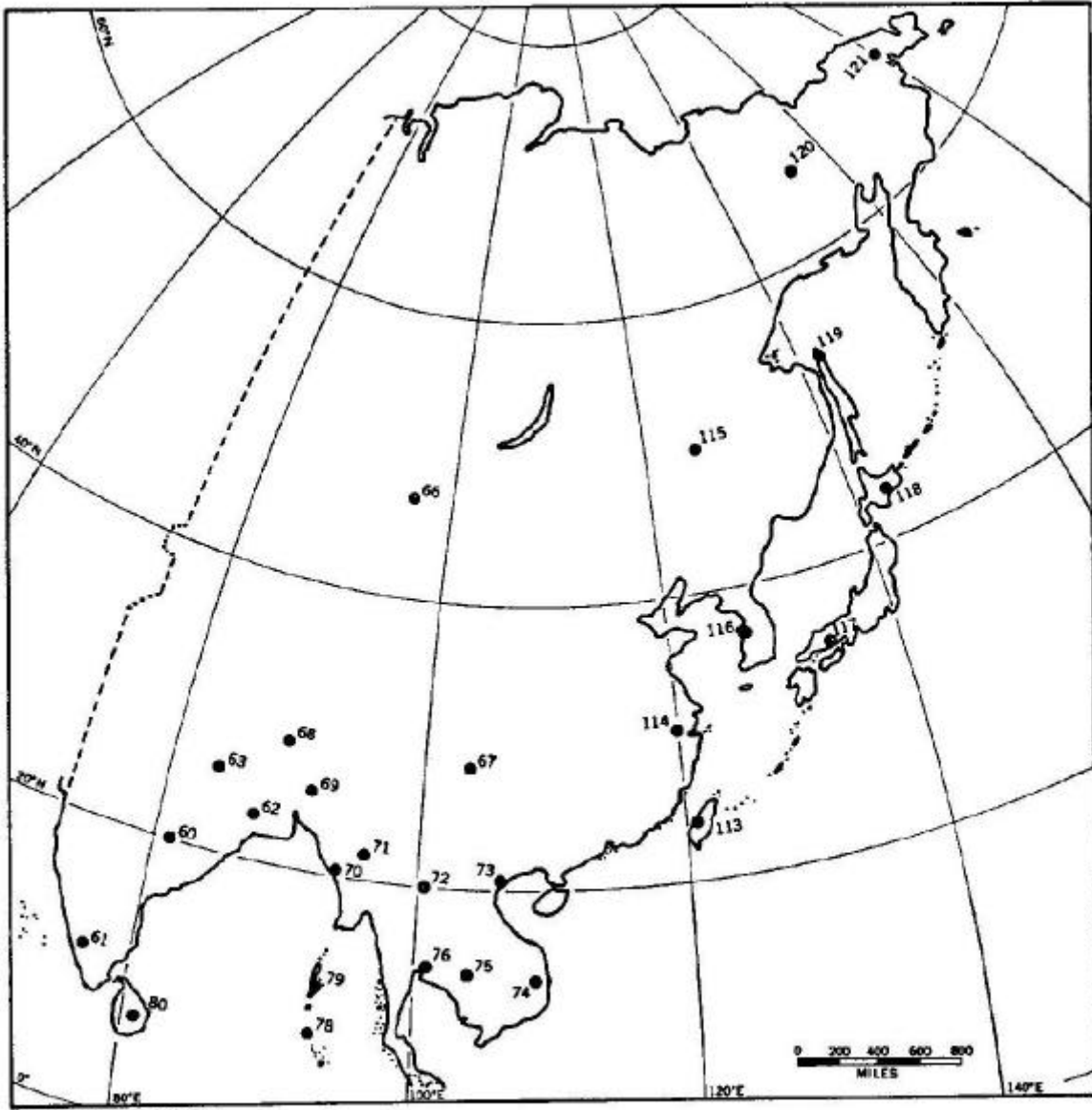


This map locates thirteen of the sample societies of the Circum-Mediterranean region (C) and five of those in East Eurasia (E).



60 Gond	1938	71 Burmese	1965	113 Atayal	1930
61 Toda	1900	72 Lamet	1940	114 Chinese	1936
62 Santal	1940	73 Vietnamese	1930	115 Manchu	1915
63 Uttar Pradesh	1945	74 Rhade	1962	116 Koreans	1947
66 Khalka Mongols	1920	75 Khmer	1292	117 Japanese	1950
67 Lolo	1910	76 Siamese	1955	118 Ainu	1880
68 Lepcha	1937	78 Nicobarese	1870	119 Gilyak	1890
69 Garo	1955	79 Andamanese	1860	120 Yukaghir	1850
70 Lakher	1930	80 Vedda	1860	121 Chukchee	1900

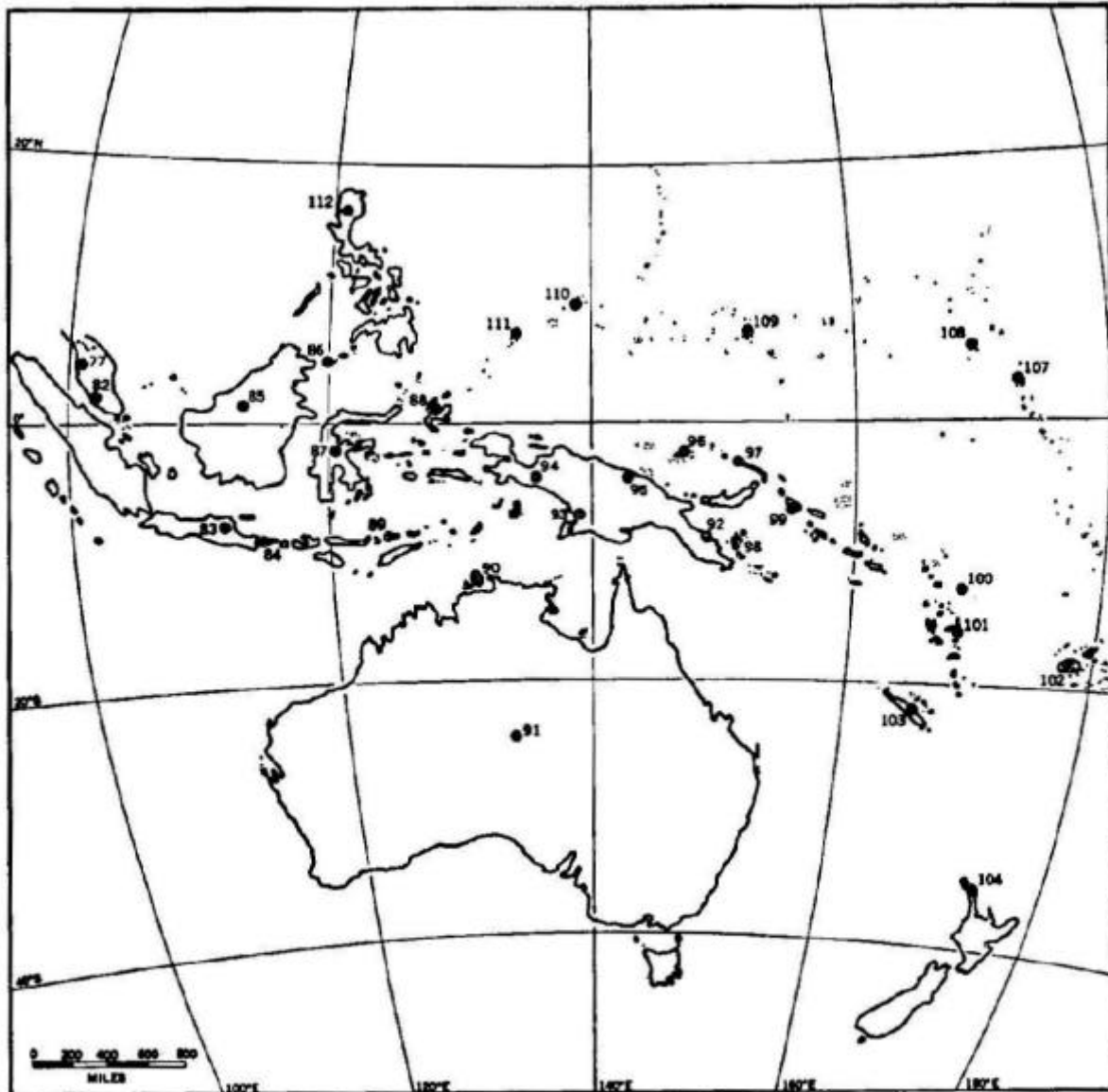
MAP 3: EAST EURASIA



This map locates 25 of the sample societies from the world region of East Eurasia (E) and one from the Insular Pacific (I).

77 Semang	1925	92 Orokaiva	1925	103 Ajie	1845
82 Negri Sembilan	1958	93 Kimam	1960	104 Maori	1820
83 Javanese	1954	94 Kapauku	1955	105 Marquesans	1800
84 Balinese	1958	95 Kwoma	1960	106 Western Samoans	1829
85 Iban	1950	96 Manus	1937	107 Gilbertese	1890
86 Badjau	1963	97 New Ireland	1930	108 Marshallese	1900
87 Toradja	1910	98 Trobrianders	1914	109 Trukese	1947
88 Tobelorese	1900	99 Siuai	1939	110 Yapese	1910
89 Alorese	1938	100 Tikopia	1930	111 Palauans	1947
90 Tiwi	1929	101 Pentecost	1953	112 Ifugao	1910
91 Aranda	1896	102 Mbau Fijians	1840		

MAP 4: INSULAR PACIFIC

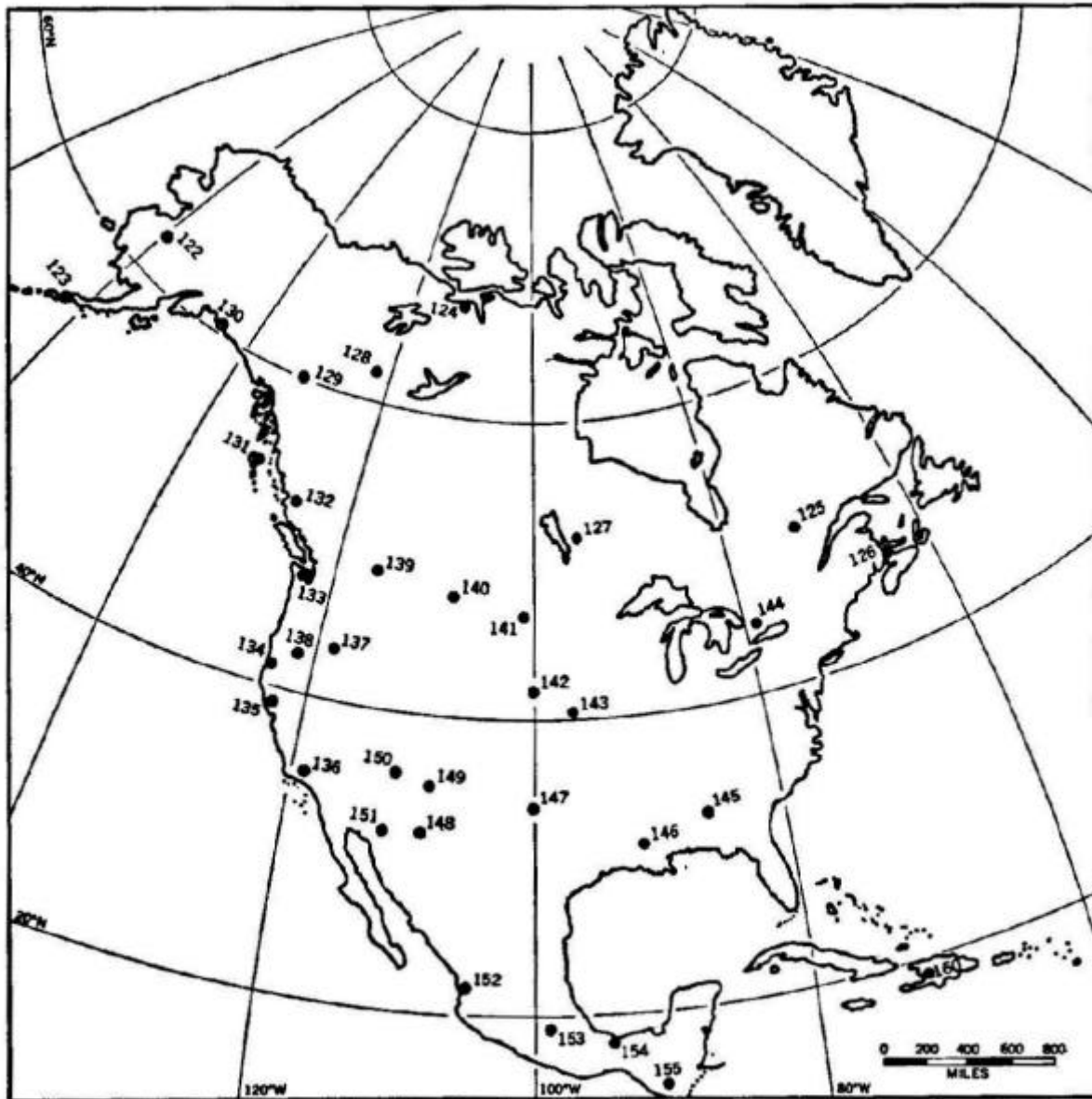


Admiral Equal-Area Projection

This map locates two of the sample societies from the East Eurasian region (E) and 28 of those from the Insular Pacific region (I).

122 Ingalik	1885	134 Yurok	1850	146 Natchez	1718
123 Aleut	1800	135 Pomo (Eastern)	1850	147 Comanche	1870
124 Copper Eskimo	1915	136 Yokuts (Lake)	1850	148 Chiricahua	1870
125 Montagnais	1910	137 Paiute (North.)	1870	149 Zuni	1880
126 Micmac	1650	138 Klamath	1860	150 Havasupai	1918
127 Sauteaux	1930	139 Kutenai	1890	151 Papago	1910
128 Slave	1940	140 Gros Ventre	1880	152 Huichol	1890
129 Kaska	1900	141 Hidatsa	1836	153 Aztec	1520
130 Eyak	1890	142 Pawnee	1867	154 Popoluca	1940
131 Haida	1875	143 Omaha	1860	155 Quiche	1930
132 Bellacoola	1880	144 Huron	1634	160 Haitians	1935
133 Twana	1860	145 Creek	1800		

MAP 5: NORTH AMERICA



Azimuthal Equal-Area Projection

This map locates all 33 of the sample societies of the North American region (N) and two of those from the South American region (S).

156 Miskito	1921	167 Cubeo (Tucano)	1939	177 Tupinamba	1550
157 Bribri	1917	168 Cayapa	1908	178 Botocudo	1884
158 Cuna (Tule)	1927	169 Jivaro	1920	179 Shavante	1958
159 Goajiro	1947	170 Amahuaca	1960	180 Aweikoma	1932
161 Callinago	1650	171 Inca	1530	181 Cayua	1890
162 Warrau	1935	172 Aymara	1940	182 Lengua	1889
163 Yanomamo	1965	173 Siriono	1942	183 Abipon	1750
164 Carib (Barama)	1932	174 Nambicuara	1940	184 Mapuche	1950
165 Saramacca	1928	175 Trumai	1938	185 Tehuelche	1870
166 Mundurucu	1850	176 Timbira	1915	186 Yahgan	1865

MAP 6: SOUTH AMERICA



Azimuthal Equal-Area Projection

**This map locates 30 of the 32 societies from the South American region (S).**

## MEASUREMENT OF HISTORICAL INFLUENCES

The delimitation of distinctive world areas as sampling strata can at best minimize, but cannot eliminate, the cultural similarities between neighboring areas which result from the historical influences of common origin and cultural diffusion. The most effective way of meeting Galton's objections is to measure the strength of such similarities on particular sets of variables. Where particular variables show high similarities between neighboring or historically (e.g., linguistically) related societies, correlations between variables should be examined to ascertain whether they have been inflated by the historical multiplication of similar cultural forms. Our standard sample reveals, for example, a correlation between agriculture and supra-community political organization with a coefficient of  $\phi = .457$ .<sup>6</sup>

For any such correlation, a solution to Galton's problem may be obtained if the portion of the correlation attributable to historical influences can be segregated from that due to functional parallelism.<sup>7</sup>

Our alignment of the sample societies in a continuous geo-cultural series from 1 to 186 makes possible the application of four of the solutions to Galton's problem proposed by Naroll (1961, 1964; Naroll and D'Andrade 1963)—the Linked Pair, Cluster, and Bimodal Sift methods, which depend upon alignment, and the Matched Pair method, which may utilize but does not require an alignment.

In the original Linked Pair method (Naroll 1964) each aligned society is compared to its neighbor (next on the list) with respect to a single variable, and a coefficient of similarity between neighbors is calculated for the entire sample. For the presence or absence of agriculture in our standard sample, for example, we have calculated a  $\phi$  coefficient of .351. Similarly, the correlation between neighboring pairs in the alignment with respect to the presence or absence of political integration transcending the community level yields a  $\phi$  coefficient of .316, with the probability that this is due to chance being  $< .0001$  or one in ten thousand.

---

<sup>6</sup>  $\phi = \text{Chi-squared} / \text{vn}$ .

<sup>7</sup> Duplication of cases in a sample may inflate or deflate correlations, or leave them unaffected, depending of whether there is a positive-, negative-, or neutral-correlation bias in the cases that are duplicated. If two correlated traits diffuse together, for example, to give a positive-correlation bias in cases duplicated, then the excess part of the correlation is due to the historical co-diffusion. Galton's problem of nonindependence, however, is also that, whatever the bias or lack of such bias, the effective sample size needs to be deflated in any case to take into account the lesser number of independent cases. Thus, while a correlation might be correctly estimated if the duplicates are unbiased, tests of the significance of the correlations that depart from zero will be exaggerated.

Given the high correlations between alignment neighbors on these two variables, it is likely that a significant portion of the correlation between agriculture and supra-community sovereignty ( $\phi = .457$ ) is attributable to an historical tendency for these two factors to diffuse or spread together (and likewise for that between the absence of agriculture and local political autonomy). The Matched Pair method of Naroll and D'Andrade (1963) can be used to estimate the relative strengths of the hypothesis of historical influence versus that of functional integration. This test, however, merely indicates whether one hypothesis is superior, equal, or inferior to the other. In the case of agriculture and supra-community sovereignty, the two hypotheses come out equally. The obvious weakness of this method is that it does not provide an estimate of what the functional correlation between the two variables would be if historical influences could be factored out; one does not know whether the residual functional correlation would be statistically significant or not. One of the solutions of Naroll and D'Andrade (1963), which would answer this question, is the Interval Sift method, in

#### STANDARD CROSS-CULTURAL SAMPLE 349

which the independent variable (e.g., economic type) is examined to ascertain the size of the "patches" of greater-than-random similarities (reflecting historical influences) between contiguous societies and to choose a subsample from representatives of distinct patches.<sup>8</sup> The present authors have found it useful to examine successively more distant pairs along the alignment to determine at what interval the similarities between pairs fall within the acceptable probabilities of a random model. This interval then becomes the criterion of the size of the "patches" of historical relatedness from which a subsample can be drawn which is free of noticeable historical effects.

---

<sup>8</sup> The average  $\phi$  correlation between agriculture and political levels using the interval sift method, however, is .527, so that the historical component of this correlation, compared to the  $\phi$  of .457 between the raw variables, is a negative-correlation bias. This might indicate that these two variables are functionally correlated but tend to diffuse independently. Galton's problem is not as simple as some investigators have assumed, namely, that nonindependence automatically inflates correlations. The take-home message is twofold: (1) the nature of the historical and diffusional aspects of Galton's Problem require separate investigation, and (2) whenever similarities among societies cluster spatially, linguistically or by other measures, such as network of linkages, it is safer to deflate sample size when calculating statistical significance of single correlations, or when comparing replication of multiple correlations. For further examination of network autocorrelation as a solution to Galton's problem, see Dow, White and Burton (1982), Dow, Burton, White and Reitz (1984), and Dow, Burton and White (1982).

In developing this modification of the Linked Pair method, we have employed an index of similarity between pairs instead of using Naroll's technique of calculating the correlation between pairs on a single attribute. This index of similarity can yield a valid measure of average similarities within regions, whereas the correlation method is unsuited for such comparison; for example, if agriculture is the dominant economic mode in a region, a correlation of the presence of agriculture between neighbors does not reflect this homogeneity. In constructing our index of similarity we have utilized the information presented in Appendix A on linguistic affiliation, subsistence economy, level of political integration, and rule of descent, and have assigned weights of 0, 1, or 2 to likenesses in each of these categories, as follows:

- 0 for essential identity, e.g., for membership in the same linguistic subfamily, for having economies of the same major type and subtype, for exhibiting the same number of levels of political integration, and for adhering to an identical rule of descent.
- 1 for partial similarity, e.g., for membership in different subfamilies of the same linguistic family, for having different subtypes of an agricultural or other major type of economy, for exhibiting a difference of only one in the number of levels of political integration, and for such relatively minor differences in descent as those between double and either matrilineal or patrilineal descent.
- 2 for maximal dissimilarity, e.g., for membership in independent linguistic families, for having different major types of subsistence economy, for exhibiting a difference of two or more in the number of levels of political integration, and for adhering to entirely different rules of descent.

Given the frequencies of each attribute for each of the four variables, a random model for the percentage of 0's, 1's, and 2's in a pairwise sample can be calculated for each variable by multiplying the co-occurrence probabilities of all possible permutations. These calculations revealed the following expected frequencies for the random model: 1.94 (standard error .05) for language, 1.32 (s.e. .09) for subsistence economy, 1.06 (s.e. .08), for political integration, and 1.48 (s.e. .07) for descent, or a total of 5.80 (s.e. .20).

The actual observed similarities between adjacent pairs in the alignment show a predictably marked divergence from the random model— $p < .01$  for political integration,  $p < .00005$  for economy and rule of descent,  $p$

< .00000001 for language. The degree of similarity is strong between adjacent pairs, as might be expected in practically any cross-cultural sample.

### 350 ETHNOLOGY

The alignment also makes it possible to evaluate successively more distant pairs (i.e., at intervals of 2, 3, etc.) to ascertain at precisely what interval, for a given variable, historical influences cease to operate, i.e., at that where the pairwise similarities are reduced to values within the expectations of the random model. Accordingly intervals of 2 and 3 were tested for the entire sample, and intervals of 4, 5, 7, 9, 10, 20, 30, and 100 for sub-samples of 10 to 20 per cent, with the results shown in Table 6.

Table 6 shows that, for the variables of political integration and subsistence economy, marked differences from randomness decrease linearly to within 1.96 standard deviations of the expected random mean [this is just the reverse of the  $p < .05$  significance test of departure from the null hypothesis, since we are here interested in the first acceptable approximation to randomness] at the intervals of 3 and 4, respectively, whereas for language and descent the decrease in similarity, also approximately linear, continues to successively

**TABLE 6. OBSERVED FREQUENCIES IN THE MEASURE OF SIMILARITY BETWEEN LINKED PAIRS AT VARIOUS INTERVALS**

Intervals	Language	Economy	Pol.Integ.	Descent	Total
1(adjacent)	1.37	.87	.84	1.10	4.18
2(alternate)	1.50	.97	.84	1.07	4.38
3	1.59	1.10	.98*	1.21	4.88
4	1.50	1.20*	1.08	1.30	5.08
5	1.58	1.26	.84	1.26	4.94
7	1.78	1.30	.81	1.27	5.16
<b>9</b>	<b>1.89*</b>	<b>1.45</b>	<b>1.05<sup>9</sup></b>	<b>1.33</b>	<b>5.72*</b>
10	1.89	1.44	1.10	1.50*	5.93
20	1.94	1.59	1.06	1.29	5.88
30	2.00	1.55	.90	1.45	5.90
100	2.00	1.48	1.38	1.30	6.16
Random	1.94	1.32	1.06	1.48	5.80

<sup>9</sup> Reconvergence to a random model; row marked to indicate that sample sizes of 20 approach nonindependence.



\* First approximation to the random model where  $p > .05$  that the value could have come from a random distribution.

more distant pairs, with differences from the random model that are statistically significant ( $p < .01$ ) up to the interval of 9. (Political integration, which deviates from randomness at the intervals of 5 and 7, returns to a second approximation thereof at the same interval of 9). This means that, for sampling considerations, the construction of an Interval Sift sample (Naroll 1961) entirely free from historical influences on language, social organization, and possibly also political complexity would require a world subsample of no more than about 20 societies (one-ninth of 186)—a number too small to yield statistically reliable correlations.<sup>10</sup> Any larger sample would presumably yield correlations between the independent and dependent variables that would be in at least some measure inflated by historical influences.

Treating economic type as the independent variable and political type as the dependent variable, an Interval Sift sample presumably free of historical influences affecting the independent variable could include all societies in

### 352 ETHNOLOGY

substitute societies can always be assigned to a specific area or province in the sample (Murdock 1968).

The superiority of our standard world sample in minimizing historical influences can be demonstrated by direct comparison with other samples, when these are arranged in the same geographical alignment. For the 220 usable societies of the HRAF sample, selected because of its nearly equivalent size, calculation of the similarities between neighboring pairs on the alignment showed a total index of similarity of 3.28, as compared with 4.18 for our standard sample and 5.80 for purely random comparisons (see Table 6). This substantial reduction in historical similarities in our sample, as compared to that of HRAF, was greatest in language, followed by descent, economy, and political integration in this order. From the magnitude of these differences it is estimated that the true value of a functional correlation, such as in our example of that between agriculture and supra-community

---

<sup>10</sup> Such correlations are unbiased estimates of the correlation in the total sample unless the two correlated traits diffuse together, but they will have low statistical significance unless averaged across different subsamples. See Dow, White and Burton (1982), Dow, Burton, White and Reitz (1984), and Dow, Burton and White (1982).

sovereignty, might frequently be doubled for the HRAF sample owing to its much stronger reflection of historical influences.

### CONCLUSION

Cross-cultural researchers who use our standard sample (or a close approximation thereto) in future studies can reap a number of advantages from so doing:

1. From our pinpointing specifications they can make certain that all elements compared actually coexisted in the same subgroup at the same point in time;

2. They can intercorrelate their findings with those of other users of the sample, including the Cross-Cultural Cumulative Coding Center at the University of Pittsburgh, which is currently engaged in coding information on all the societies of the sample and a substantial number of alternates for three sets of codes—on infancy and the transition to childhood, on subsistence economy and the local community, and on settlement pattern and community organization;

3. They need not use all the 186 societies of the standard sample but may conveniently work with smaller subsamples of 83 (half of the standard sample consisting of every alternate society in it) or 62 (a third of the standard sample consisting of every third society), either being equally representative of the world's known and well described cultures as well as exhibiting slightly less contamination from historical influences;

4. By using the same geo-cultural alignment they can employ the same methods of measuring or controlling for historical influences;

5. They may for any good reason substitute other societies from the same distinctive areas or sampling provinces without sacrificing any of the advantages except possibly that of intercorrelation with the results of other studies.

## **Appendix A: Specific Data on Each Sample Society**

Data omitted are available on the web

at <http://eclectic.ss.uci.edu/~drwhite/worldcul/Scs34.htm> ([go there](#)).

BIBLIOGRAPHY

- Naroll, R. 1961. Two Solutions to Galton's Problem. **Philosophy of Science** 28:15-39.
- 1964. A Fifth Solution to Galton's Problem. **American Anthropologist** 66:863-867.
- Naroll, R. and R. G. D'Andrade. 1963. Two Further Solutions to Galton's Problem. **American Anthropologist** 65: 1053-1067.
- Ogburn, W. F. 1922. **Social Change**. New York.
- Sumner, W. G. 1906. **Folkways**. Boston.

ADDENDA 2002

- Dow, M., M. L. Burton, D.R. White. 1982. Network Autocorrelation: A Simulation Study of a Foundational Problem in Regression and Survey Research. **Social Networks** 4:169-200.
- Dow, M., M.L. Burton, D.R. White, K. Reitz. 1984. Galton's Problem as Network Autocorrelation. **American Ethnologist** 11:754-770.
- Dow, M., D. R. White, M.L. Burton. 1982. Multivariate Modeling with Interdependent Network Data. **Behavior Science Research** 17:216-245.
- Lagacé, Robert O. 1979. The HRAF Probability Sample: Retrospect and Prospect. **Behavior Science Research**, 14:211-229.
- Murdock, George P. 1967. *Ethnographic Atlas*. University of Pittsburgh Press. Also *Ethnology* 6:109-236.
- Naroll, Raoul. 1967. The Proposed HRAF Probability Sample. **Behavior Science Notes** 2:70-80.
- White, D. R., M. Dow, M. L. Burton. 1981. Sexual Division of Labor in African Agriculture: A Network Autocorrelation Analysis. **American Anthropologist** 83:824-849.

-- -- not cited

Murdock, George P. 1949. **Social Structure**. New York: Macmillan.

Murdock, George P. 1981. **Atlas of World Cultures**. Pittsburgh: University of Pittsburgh Press.

Murdock, George P. 1983. **Outline of World Cultures**, 6th ed. New Haven: Human Relations Area Files.

Murdock, George P., Clellan S. Ford, Alfred E. Hudson, Raymond Kennedy, Leo W. Simmons, John W. M. Whiting. 5th revised edition. 1987. **Outline of Cultural Materials**. New Haven: Human Relations Area Files.

White, D. R., and R. Pesner. 1983. Internal Replication and the Systems Concept in Non-Experimental Research. **Behavior Science Research** 18:26-44.

White, D. R., R. Pesner, and K. Reitz. 1983. "An Exact Significance Test for Three-Way Interaction. **Behavior Science Research** 18:103-122.