2004 World Cultures 15(2):153-170

Does Mr. Galton Still Have a Problem? Autocorrelation in the Standard Cross-Cultural Sample

E Anthon Eff
Department of Economics and Finance, Middle Tennessee State University, Murfreesboro, TN 37132; eaeff@mtsu.edu

This paper examines the Standard Cross-Cultural Sample (SCCS), extracted by George P. Murdock and Douglas R. White from the Ethnographic Atlas in an effort to produce independent observations, so that statistical analysis would not be biased by "Galton's Problem." Spatial statistics are used to test for spatial autocorrelation and "phylogenetic" autocorrelation (using language phyla to signify degree of relatedness) across the 186 cultures in the sample. The results suggest that the cultures are not completely independent, and that it would be prudent to test for spatial and phylogenetic autocorrelation when conducting regression analyses with the Standard Cross-Cultural Sample.

1. INTRODUCTION

In 1889, Edward Tylor presented what was to become the seminal paper in statistical cross-cultural analysis, before a panel at the Royal Anthropological Institute. Sitting on the panel was Sir Francis Galton, the statistician and eugenicist. Tylor compiled information on institutions of marriage and descent for 350 cultures and examined the correlations among these institutions. The results showed that certain institutions were associated with each other far more often than chance would imply; Tylor interpreted these results as indications of a general evolutionary sequence, in which institutions changed focus from the maternal line to the paternal line. Galton disagreed, pointing out that similarity between cultures could be due to borrowing, could be due to common descent, or could be due to evolutionary development; he maintained that without controlling for borrowing and common descent one cannot make valid inferences regarding evolutionary development. In the literature, Galton's critique has become the eponymous "Galton's Problem." (Gillies 2000; Stocking 1968: 175).

Galton's problem appears in a peculiar form in regression analysis. The statistical technique requires that the disturbance term in the estimated model have certain properties, one of which is that the disturbances not be correlated with each other. Violation of this property causes the estimated standard errors of the coefficients to be biased, so that one cannot trust the t-statistics, and one therefore cannot make hypothesis tests regarding the estimated coefficients (Kennedy 1998). Hence, cross-cultural analysis employing the method of regression must take seriously the prospect that the disturbance terms may be correlated. The likely sources of such correlation are exactly the sources mentioned by Sir Francis Galton: individual cultures are probably related, either by descent or via cultural diffusion.

Galton's Problem has received much attention from practitioners of cross-cultural analysis. George Peter Murdock attempted to tackle the problem by developing a sample of cultures relatively independent from each other—i.e., with relatively weak phylogenetic and cultural diffusion relationships. Murdock began with the twelve hundred or so peoples in his *Ethnographic Atlas* (Murdock 1957), dividing them into roughly 200 "sampling provinces" of closely related cultures. Working with Douglas R. White, Murdock chose one particularly well-documented culture from each sampling province, to create the Standard Cross-Cultural Sample (SCCS) (Murdock and White 1969). The number of cultures is large and varied enough to provide a sound basis for statistical analysis; the sample includes 186 cultures, ranging from contemporary hunter gatherers (e.g., the Mbuti), to early historic states (e.g., the Romans), to contemporary industrial peoples (e.g., the Russians) (Silverman and Messinger 1999; Mace and Pagel 1994). The names of the cultures, their locations, their language families, and their levels of cultural complexity are listed in the appendix.[1]

Scholars engaging in statistical cross-cultural analysis are encouraged to use the set of cultures in the SCCS, since each new study adds to the number of coded variables capable of being used with already existing variables. The electronic journal World Cultures functions as the repository of the SCCS, archiving the now nearly 2000 coded variables and publishing a number of papers on cross-cultural methodology.

Murdock and White performed an important service in selecting the SCCS, but one can readily understand that many cultures in the sample do have links to other cultures in the sample, so that statistical independence is not fully realized. For example, the sample contains 13 cultures speaking Bantoid languages and four cultures speaking Cushitic languages. This suggests that the sample might mitigate Galton's Problem, but not eliminate it. A long line of empirical work (Naroll 1965; Loftin 1972; Loftin and Ward 1983; Mace and Pagel 1994; Hays 1998) recognizes the need to test for and correct phylogenetic relationships when performing statistical analysis with cross-cultural data sets. In the remainder of the paper, Moran's I—a statistic drawn from spatial econometrics—is used to test for the degree of relatedness among cultures for each of the variables in the SCCS. The test is performed both for phylogenetic relatedness and for spatial proximity. The results indicate the degree to which variables in the sample are affected by autocorrelation from the two sources. In the final portion of the paper, an example is given of how Moran's I can be used in regression analysis. The example estimates a model of female contribution to subsistence, and then evaluates the errors to determine if phylogenetic or spatial autocorrelation creates a problem.

2. TESTING FOR AUTOCORRELATION IN THE SCCS

The statistic used here is Moran's I (Odland 1988; Anselin 1988):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})} \quad (1)$$

Where w_{ij} is a weight representing the degree of relatedness between location i and j (greater relatedness implies a higher weight); n is the number of locations; x_i is the value of a variable at location i and x_j is the value of the same variable at location j.

The statistic is in the class of correlation statistics, since the term on the right is the ratio of a covariance to a variance (Odland 1988). Intuitively, it differs from the usual correlation coefficient in that a correlation coefficient compares the values of two variables at each location, while Moran's I compares the value of a single variable for each pair of locations arrayed according to degree of relatedness.

Because it is in the class of correlation coefficients, one can calculate a variance for a Moran's I. This then allows one to conduct hypothesis tests, by setting the difference between the Moran's I and its hypothesized value over the standard error of the Moran's I. The resulting t-distribution can be used to reject or maintain a null hypothesis that there is no autocorrelation (Odland 1988; Anselin 1988).

Three different sets of weights were constructed. The first set represents the spatial distance: using latitude and longitude, the great circle distance was calculated between each pair of cultures. The squared inverse of this distance was then employed as the weight. If two locations are more likely to have similar features in a cultural trait when they are physically close, then the Moran's I will be positive and one can reject the null hypothesis of no spatial autocorrelation. Spatial distance should produce significant autocorrelation in cases where the trait is diffused by borrowing.

The second set of weights represents phylogenetic relationships. Some genetic data have been compiled for human populations, most notably by Cavalli-Sforza, Menozzi, and Piazza (1994), and these authors have noted that the genetic distance between any pair of populations matches fairly closely the linguistic distance between those populations. There are no data for genetic distance for the set of cultures in the SCCS, but the data set does contain the language continent, phylum, and family, for each culture. Weights are given a value of 0 if the pair of cultures are in different language continents, a value of 1 if in the same language continent, a value of 2 if in the same language phylum, and a value of 3 if in the same language family. Linguistic distance should produce significant autocorrelation in cases where the trait is transmitted through descent.

The third set of weights represents cultural complexity. In the initial discussion, it was mentioned that Edward Tylor initiated statistical cross-cultural analysis in an effort to identify characteristics of cultural evolution. Many social scientists using the SCCS maintain this particular interest. If a cultural trait is associated with larger and more differentiated societies, it could well be that that trait is called forth through an adaptive process of mutual causation, where changes in one part of the social structure elicit changes in other parts, in order to solve adaptive problems. This kind of functional, evolutionary perspective is at least as old as Herbert Spencer (1897), and is of interest in the present context because it depicts a way in which a cultural trait may be acquired that is distinct from both descent and borrowing. The index of cultural complexity (fashioned from variables variables 149-158 in the SCCS) goes from 1 (least complex) to 20 (most complex). The weights are the inverse of the squared difference in cultural complexity between each pair of cultures. If two locations are more likely to have similar features in a cultural trait when they are at similar levels of cultural complexity, then the Moran's I will be positive and one can reject the null hypothesis of no autocorrelation in the dimension of cultural complexity. Significant autocorrelation should occur in cases where the trait is elicited through cultural evolution. The societies in the SCCS and their codes for location, linguistic affiliation, and cultural complexity are listed in appendix 1.

Table 1 presents a summary of the results of the Moran's I test on 1700 variables in the SCCS. The three columns at the right represent the percent of variables for which the specified type of autocorrelation was present (using a 0.10 size of test). The bottom row (next page) shows that about 44% of the variables exhibit spatial autocorrelation, 43% exhibit linguistic autocorrelation, and 44% exhibit autocorrelation in the dimension of cultural complexity. The results show that relations of borrowing and descent are present in the data, and that the degree to which they are present is quite strong: as strong as the degree to which evolutionary relations are present.

Table 1: Percent of Variables with Significant Autocorrelation, by Topic

SCCS Variables	Number of Variables	Topic	Distance	Language	Cultural Complexity
1-22	22	Subsistence Economy and Supportive Practices	91%	95%	91%
23-60	38	Infancy and Early Childhood	45%	55%	47%
61-80	20	Settlement Patterns and Community Organization	65%	65%	65%
81-98	18	Political Organization	67%	72%	78%
99-148	48	Division of Labor	67%	60%	42%
149-158	10	Cultural Complexity	100%	100%	100%
159-178	20	Sexual Attitudes and Practices	30%	15%	25%
179-199	21	Climate Data from Weather Stations	95%	86%	67%
200-292	92	Ethnographic Atlas	71%	74%	68%
293-336	44	Traits Inculcated in Childhood	45%	55%	39%
337-480	144	Agents and Techniques of Child Training	35%	39%	19%
481-528	42	Parental Acceptance-Rejection and Parental Control	12%	10%	5%
529-560	32	Adolescent Initiation Ceremonies	100%	97%	47%
561-575	15	Reproductive Rituals	93%	100%	47%

576-636	60	The Relative Status of Women	48%	32%	22%
637-644	8	Kin Term Patterns	63%	50%	25%
645-656	12	Cultural Theories of Illness	42%	42%	42%
657-679	23	Female Power and Male Dominance	57%	26%	22%
680-738	59	Female Status: Independent Variables	66%	61%	56%
739-755	17	Husband-Wife Relationships	41%	41%	24%
756-797	42	Political Decision Making and Conflict	48%	45%	57%
798-813	16	Data Quality Control Variables for Child Training	44%	25%	25%
814-826	13	Sexual Division of Labor Revisited	92%	92%	92%
827-832	6	Adolescent Sexual Behavior	50%	0%	83%
833-850	7	Standard Cross-Cultural Sample	57%	71%	71%
854-859	6	Climate and Subsistence	100%	100%	100%
860-879	20	Polygyny: Form and Frequency	95%	90%	90%
879-884	6	Magico-Religious Practitioners	50%	67%	83%
885-890	6	Female Contribution to Subsistence	83%	67%	83%
891-916	26	The Nature of Warfare	54%	50%	73%
917-920	4	Slavery and Social Death	100%	100%	100%
921-930	10	Agricultural Potentials	80%	80%	60%
931-985	55	Varieties of Sexual Experience	31%	33%	29%
986-1005	20	Enculturative Continuity and Importance of Caretakers	10%	0%	5%
1006-1114	67	Historical Analysis of Subsistence Change	33%	7%	22%
1122-1122	1	Population Codes	100%	100%	100%
1123-1131	9	Type of Agriculture	56%	89%	44%
1132-1132	1	State Organization	100%	100%	100%
1133-1135	3	Despotism and Harem Size	100%	67%	67%
1136-1178	43	Divorce	12%	40%	23%
1188-1189	2	Evil Eye	100%	100%	100%
1190-1225	29	Kin Avoidance	7%	10%	14%
1248-1252	4	Female Beauty and Adolescent Sexuality Codes	50%	50%	25%
1253-1260	8	Pathogen Stress Cross-Culturally: Codes	100%	100%	75%
1261-1270	10	Starvation and Famine Among SCCS Societies: Codes	50%	50%	50%
1271-1305	35	Household Division Of Work: primary crop 1	9%	37%	71%
1306-1341	36	Household Division Of Work: primary crop 2	28%	58%	92%
1342-1366	24	Household Division Of Work: primary crop 3	33%	38%	67%
1367-1401	29	Household Division Of Work: secondary crop 1	7%	17%	83%
1402-1437	36	Household Division Of Work: secondary crop 2	17%	6%	92%
1438-1462	24	Household Division Of Work: secondary crop 3	21%	21%	67%
1463-1490	28	Household Division Of Work: small animals	54%	57%	46%
1491-1521	31	Household Division Of Work: large animals	65%	29%	68%
1522-1557	36	Household Division Of Work: wage, trade, gathering	33%	17%	36%
1558-1591	34	Household Division Of Work: hunting, child care, housekeeping	35%	15%	21%
1592-1614	18	Household Division Of Work: cooking, fire tending	11%	0%	28%
1615-1647	32	Household Division Of Work: gather fuel, carry burdens, ca	19%	13%	9%

1648-1691	44	Warfare, Aggression, and Resource Problems	7%	18%	9%
1692-1709	17	Scarification, Pathogen Load and Biome	71%	76%	59%
1710-1713	4	Sleeping Arrangements of Children & Adolescents	75%	75%	50%
1714-1747	34	CONAN: Code-Text Data-Base, Part I	29%	56%	68%
1748-1780	33	CONAN: Code-Text Data-Base, Part II	27%	33%	27%
1781-1805	25	Codes on Gossip	12%	24%	8%
1-1805	1700	Total SCCS	44%	43%	44%

3. AUTOCORRELATION TESTS IN EXPLORATORY DATA ANALYSIS

The various species of autocorrelation—spatial, linguistic, and "evolutionary"—are present in various degrees in each of the different topics listed in Table 2, and the study of these autocorrelation results can provide useful insights as part of an exploratory data analysis. For example, Schlegel and Barry's (1979) study on adolescent initiation rituals contains 26 variables—13 pertinent to boys and 13 pertinent to girls—describing the experience of adolescents in each culture during initiation. The first column in Table 3 describes the variable, and the statement in parentheses give some sense of what an increase in the variable's value implies. Thus, the first row indicates that an increasing value for "Occurrence" implies that adolescent initiation rituals are present. The table presents the p-values for the autocorrelation t-statistic in each of the three dimensions (distance, language, cultural complexity); p-values for girls are presented in the first three columns, and p-values for boys in the next three.

The distance p-value and the language p-value are below 0.05 in all instances—allowing us to reject the null hypothesis of no autocorrelation. Cultural complexity, however, is a different story: girls clearly have autocorrelation in the dimension of cultural complexity while boys just as clearly do not. One can thus make the preliminary inference that relations of borrowing and descent account for the presence of particular adolescent initiation rituals for both sexes; "evolutionary" relations account for the presence of adolescent initiation rituals for girls, but not for boys. As part of an exploratory data analysis this finding could then trigger a more rigorous and detailed effort at uncovering the features and sources of this "evolutionary" sex difference.

Table 2. Autocorrelation (three types) for Variables in Schlegel and Barry's (1979) Study on Adolescent Initiation

Variable	Girls			Boys		
	Distance	Language	Cultural Complexity	Distance	Language	Cultural Complexity
V529-V530 Occurrence (↑=present)	0%	0%	0%	0%	0%	28%
V531-V532 Age (↑=older)	0%	0%	0%	0%	0%	73%
V533-V534 Number of Concurrent Initiates (↑=group larger)	0%	0%	1%	0%	0%	49%
V535-V536 Duration of Ceremony (↑=longer)	0%	0%	0%	0%	0%	52%
V537-V538 Number of Participants (↑= larger)	0%	0%	2%	0%	0%	86%
V539-V540 Sexes of Participants (↑=more sex segregation)	0%	0%	0%	0%	0%	65%
V541-V542 Primary Physical Components (↑=more intense pain)	0%	0%	2%	0%	0%	37%
V543-V544 Secondary Physical Components (↑=more intense pain)	0%	0%	1%	0%	0%	89%
V545-V546 Primary Cognitive or Performance Components (↑=intensity of ordeal increases)	0%	0%	0%	0%	0%	52%
V547-V548 Secondary Cognitive or Performance Components (↑=intensity of ordeal increases)	0%	0%	0%	0%	0%	64%
V549-V550 Primary Emic Interpretations (↑=intensity of ordeal increases)	0%	0%	0%	0%	0%	77%
V551-V552 Secondary Emic Interpretations (↑=intensity of ordeal increases)	0%	0%	0%	1%	0%	77%
V553-V554 Tertiary Emic Interpretations (↑=intensity of ordeal increases)	0%	2%	0%	1%	1%	51%

4. AUTOCORRELATION TESTS IN REGRESSION ANALYSIS

In the building of regression models, one must ensure that the error term has certain properties. One of those properties is that the residuals not be correlated with each other i.e., that there be no autocorrelation. The presence of autocorrelation in the regression residuals causes the estimated standard errors to be biased, and thus invalidates the t-statistics and any hypothesis tests conducted with those t-statistics. With a few modifications (Odland 1988;

Anselin 1988), the Moran's I can used to test the regression residuals for the presence of autocorrelation. The important dimensions to consider in a cross-cultural data set would be the dimensions of distance and descent.

Table 4 presents the results of an ordinary least squares regression. The model attempts to explain variations in the average female contribution to subsistence across cultures. Interpreting these results in the usual way, one would say that the average female contribution to subsistence is higher in societies where fishing and hunting are less important food sources, where total pathogen stress is lower, where rainfall is lower, where polygamy is more common, where mothers spend more time with young sons, where land transport relies more on human power, where political integration occurs over more levels, and where there are fewer evidences of social stratification.

Table 4. Regression Explaining SCCS Variable #826 (Average Female Contribution to Subsistence)

Variable	Label	Parameter Estimate	Standard Error	t-value	Pr > |t|	VIF
Intercept	Intercept	46.00715	9.49409	4.85	<.0001	0
V816	Importance Fishing %	-0.19866	0.06649	-2.99	0.0032	1.41144
V817	Importance Hunting %	-0.42973	0.08083	-5.32	<.0001	1.56621
V1260	Total Pathogen Stress	-1.24132	0.39485	-3.14	0.002	2.03294
V855	Niche Rainfall	-2.28053	0.65284	-3.49	0.0006	1.4644
V79	Polygamy	2.95054	1.75657	1.68	0.0949	1.37957
V353	Sex of Parent in Residence: Early Boy	5.4094	1.77529	3.05	0.0027	1.43635
V150	Fixity of Residence	-1.41896	0.91557	-1.55	0.1231	1.82325
V154	Land Transport	-2.35077	1.1751	-2.00	0.047	1.67645
V157	Political Integration	2.56485	1.33416	1.92	0.0562	2.19442
V158	Social Stratification	-2.45671	1.1415	-2.15	0.0328	2.46027
	R-Square	0.3121				
	Adj R-Sq	0.2714				
	F Value	7.67				
	Pr > F	<.0001				
	p-value on Moran's I: Distance	0.0650				
	p-value on Moran's I: Language	0.0102				

This usual interpretation, however, would be incorrect. The p-values for the Moran's I tests at the bottom of Table 4 show that there is significant autocorrelation in both the distance and descent dimension. Therefore, the reported standard errors are incorrect, and the t-statistics and resulting p-values are wrong, so that our inferences are specious.

When autocorrelation exists in a regression model, it can usually be removed by the appropriate respecification. In most cases, this requires the creation of a "lagged" dependent variable—i.e., a weighted average of the dependent variable at all other locations, where the weights indicate the degree of relationship. Estimation of regression models containing lagged dependent variables, however, must usually be done using maximum likelihood methods: ordinary least squares biases the estimated coefficients since the lagged variables typically are endogenous (Anselin 1988).

5. SUMMARY AND CONCLUSIONS

Galton's problem refers to the fact that statistical inferences from cross-cultural data must control for relations of borrowing and descent. The Standard Cross-Cultural Sample is an important effort to address Galton's problem by utilizing a sample of cultures that have relatively weak relations of borrowing and descent. The paper showed, however, that about 44% of the variables in the Standard Cross-Cultural Sample exhibit autocorrelation in the spatial and/or linguistic dimensions, indicating the transmission of cultural traits through borrowing and/or descent in these cases. This level of autocorrelation is about the same as the level of autocorrelation in the dimension of cultural complexity—indicating that the degree to which borrowing or descent affect the presence of a cultural trait is about the same as the degree to which cultural evolution affects the presence of a cultural trait. Therefore, users of the Standard Cross-Cultural Sample must not assume that the cultures are sufficiently unrelated to allow one to ignore Galton's Problem.

The paper then demonstrated how the Moran's I statistic could be used in exploratory data analysis. The example focused on characteristics of adolescent initiation rituals, and showed that autocorrelation along the dimensions of borrowing and descent were similar for boys and girls, but differed sharply along the dimension of evolutionary development: girls' initiation rituals appear to be strongly influenced by cultural evolution, while those for boys are not. This pattern could constitute the starting point for a more rigorous and detailed analysis.

Finally, the paper presented an example of a regression model employing variables from the Standard Cross-Cultural Sample. A variant of Moran's I was used to estimate autocorrelation in the model's residuals, in the dimensions of distance (i.e., borrowing) and language (i.e., descent). The residuals were significantly autocorrelated in both dimensions, rendering the results of the regression analysis invalid. One might reasonably generalize that autocorrelation is likely to be a problem with other regression analyses using the Standard Cross-Cultural Sample, and that one should therefore routinely test for the presence of autocorrelation when using these data.

6. NOTES

1. An index, produced by myself, using variables 149-158 in the SCCS.

2. For the detailed description of these variables see appendix 2, where the relevant pages from Divale (2000) are reproduced.

7. REFERENCES CITED

Anselin, L.
1988 *Spatial Econometrics: Methods and Models*. Dordrecth Academic Publishers.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza
1994 *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Divale, William
2000 Pre-coded variables for the Standard Cross-Cultural Sample from *World Cultures, Volumes I and II.*

Dow, M., M. Burton, and D. White
1982 Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks* 4:169-200.

Dow, M., M. Burton, D. White, and K. Reitz
1984 Galton's problem as network autocorrelation. *American Ethnologist* 11:754-770.

Gillies, Judith L.
2000 Cross-Cultural Analysis. *Anthropological Theories: A Guide Prepared by Students for Students*. Department of Anthropology, College of Arts and Sciences. University of Alabama (http://www.as.ua.edu/ant/Faculty/murphy/crosscut.htm).

Hayes, D. G.
1998 *The Measurement of Cultural Evolution in the Non-Literate World: Homage to Raoul Naroll*. New York: Metragram Press.

Hoy, A.
1994 The relationship between male dominance and militarism: Quantitative tests of several theories. *World Cultures* 8(2):40-57.

Kennedy, P.
1998 *A Guide to Econometrics*. 4th edition. Cambridge: MIT Press. New York: Guilford.

Loftin, C.
1972 Galton's problem as spatial autocorrelation: Comments on Ember's empirical test. *Ethnology* 11:425-435.

Loftin, C., and S. Ward
1983 A spatial autocorrelation model of the effects of population density on fertility. *American Sociological Review* 48:121-128.

Mace, R., and M. Pagel
1994 The comparative method in anthropology. *Current Anthropology* 35:549-564.

Murdock, G. P., and D. R. White
1969 A standard cross-cultural sample. *Ethnology* 9:329-369.

Naroll, R.
1965 Galton's problem: The logic of cross-cultural research. *Social Research* 32:428-451.

Odland, J.
1988 *Spatial Autocorrelation*. Newbury Park, CA: Sage Publications.

Schlegel, A., and H. Barry III
1979 Adolescent initiation ceremonies. *Ethnology* 18:199-210.
Silverman, P. and J. Messinger
1999 *Introduction to the Standard Cross-Cultural Survey Module*. SSRIC Teaching Resources Depository Standard Cross-Cultural Sample Survey. California State University, Bakersfield (http://www.csubak.edu/ssric/Modules/SCCS/SCCSMod/sccsin.htm)
Spencer, H.
1897 *The Principles of Sociology.* 3rd Edition. New York: D. Appleton.
Stocking, G. W., Jr.
1968 Edward Burnett Tylor. *International Encyclopedia of the Social Sciences* (D. Sills, ed.). New York: v16:170-177.
White, D. R., M. Burton, and M. Dow
1981 Sexual division of labor in African agriculture: A network autocorrelation analysis. *American Anthropologist* 83:824-849.

8. APPENDIX 1. SOCIETIES IN THE STANDARD CROSS-CULTURAL SAMPLE

Society	Cultural Complexity	Date	Latitude	Longitude	Region	Language Family
1 Nama (Hottentot)	8	1860	-27.500	17.000	Africa	Southern Khoisan
2 Kung (San)	3	1950	-19.833	20.580	Africa	Southern Khoisan
3 Thonga	12	1865	-25.833	32.333	Africa	Bantoid
4 Lozi	13	1900	-16.000	23.500	Africa	Bantoid
5 Mbundu	13	1890	-12.250	16.500	Africa	Bantoid
6 Suku	12	1920	-6.000	18.000	Africa	Bantoid
7 Bemba	11	1897	-10.500	30.500	Africa	Bantoid
8.Nyakyusa (Ngonde)	11	1934	-9.500	34.000	Africa	Bantoid
9 Hadza	1	1930	-3.750	35.180	Africa	Northern Khoisan
10 Luguru	11	1925	-6.833	37.667	Africa	Bantoid
11 Kikuyu	11	1920	-0.667	37.167	Africa	Bantoid
12 Ganda	15	1875	0.333	32.500	Africa	Bantoid
13 Mbuti (Pygmies)	1	1950	1.500	28.333	Africa	Bantoid
14 Nkundo (Mongo)	12	1930	-0.750	19.167	Africa	Bantoid
15 Banen	11	1935	4.667	10.800	Africa	Bantoid
16 Tiv	11	1920	7.250	9.000	Africa	Bantoid
17 lbo (Igbo)	14	1935	5.500	7.333	Africa	Kwa
18 Fon	15	1890	7.200	1.910	Africa	Kwa
19 Ashanti (Twi)	14	1895	7.000	-1.500	Africa	Kwa
20 Mende	13	1945	7.833	-12.000	Africa	Mande
21 Wolof	15	1950	13.750	-15.333	Circum-Mediterranean	Atlantic
22 Bambara	15	1902	12.500	-7.000	Africa	Mande
23 Tallensi	13	1934	10.660	-0.567	Africa	Voltaic
24 Songhai	14	1940	16.583	-1.667	Circum-Mediterranean	Songhai
25 Wodaabe Fulani	7	1951	15.000	7.500	Circum-Mediterranean	Atlantic

26 Hausa	16	1900	10.500	7.500	Circum-Mediterranean	Chadic
27 Massa	10	1910	10.500	15.500	Africa	Chadic
28 Azande	13	1905	5.083	28.250	Africa	Eastern Niger-Congo
29 Fur	14	1880	13.500	25.500	Circum-Mediterranean	Fur
30 Otoro Nuba	11	1930	11.333	30.667	Africa	Kordofanian
31 Shilluk	13	1910	9.750	31.500	Africa	Eastern Nilotic
32 Mao	10	1939	9.267	34.667	Africa	Komam
33 Kaffa	14	1905	7.267	36.500	Circum-Mediterranean	Cushitic
34 Masai	8	1900	-3.500	36.750	Africa	Eastern Nilotic
35 Konso	14	1935	5.250	37.500	Circum-Mediterranean	Cushitic
36 Somali	12	1900	9.000	47.250	Circum-Mediterranean	Cushitic
37 Amhara	15	1953	12.500	37.750	Circum-Mediterranean	Semitic
38 Bobo	12	1855	15.750	38.750	Circum-Mediterranean	Cushitic
39 Kenuzi Nubian	14	1900	23.000	38.750	Circum-Mediterranean	Nubian
40 Teda	9	1950	20.500	17.500	Circum-Mediterranean	Saharan
41 Tuareg	12	1900	23.000	6.500	Circum-Mediterranean	Berber
42 Riffians	16	1926	34.917	-3.250	Circum-Mediterranean	Berber
43 Egyptians (Fellah)	18	1950	24.750	33.000	Circum-Mediterranean	Semitic
44 Hebrews	17	-621	31.180	34.917	Circum-Mediterranean	Semitic
45 Babylonians	19	-1750	32.583	44.750	Circum-Mediterranean	Semitic
46 Rwala Bedouin	10	1913	33.250	38.500	Circum-Mediterranean	Semitic
47 Turks	18	1950	39.333	34.250	Circum-Mediterranean	Turkic
48 Gheg (Albanians)	15	1910	42.000	20.167	Circum-Mediterranean	Albanian
49 Romans	19	110	41.667	13.500	Circum-Mediterranean	Romance
50 Basques	17	1934	43.250	-1.667	Circum-Mediterranean	Basque
51 Irish	17	1932	53.500	-10.000	Circum-Mediterranean	Celtic
52 Lapps	8	1950	68.700	21.500	Circum-Mediterranean	Finno-Ugric
53 Yurak (Samoyed)	7	1894	68.000	51.500	East Eurasia	Samoyed
54 Russians	18	1955	52.667	41.333	Circum-Mediterranean	Balto-Slavic
55 Abkhaz	14	1880	43.125	40.770	Circum-Mediterranean	North Caucasian
56 Armenians	16	1843	40.000	44.500	Circum-Mediterranean	Armenian
57 Kurd	16	1951	36.500	44.500	Circum-Mediterranean	Indo-Iranian

58 Basseri	11	1958	29.000	53.000	East Eurasia	Indo-Iranian
59 West Punjabi	17	1950	32.500	74.000	East Eurasia	Indo-Iranian
60 Gond	11	1938	19.625	80.917	East Eurasia	Central Dravidian
61 Toda	8	1900	11.500	76.500	East Eurasia	Southern Dravidian
62 Santal	12	1940	23.500	87.167	East Eurasia	Munda
63 Uttar Pradesh	20	1945	25.917	83.000	East Eurasia	Indo-Iranian
64 Burusho	13	1934	36.433	74.583	East Eurasia	Burusho
65 Kazak	13	1885	42.500	75.500	East Eurasia	Turkic
66 Khalka Mongols	14	1920	47.167	96.083	East Eurasia	Mongolian
67 Lolo	14	1910	27.500	103.500	East Eurasia	Lolo-Burmese
68 Lepcha	13	1937	27.500	89.000	East Eurasia	Bodo-Naga-Kachin
69 Garo	12	1955	26.000	91.000	East Eurasia	Bodo-Naga-Kachin
70 Lakher	12	1930	22.333	93.000	East Eurasia	Naga-Kuki
71 Burmese	18	1965	22.000	95.667	East Eurasia	Lolo-Burmese
72 Lamet	8	1940	20.000	100.667	East Eurasia	Palaung-Wa
73 Vietnamese	17	1930	20.500	106.250	East Eurasia	Annam-Muong
74 Rhade	12	1962	13.000	108.000	East Eurasia	Hesperonesian
75 Khmer	17	1292	13.000	103.833	East Eurasia	Khmer
76 Siamese	20	1955	14.000	100.850	East Eurasia	Thai-Kadai
77 Semang	3	1925	5.000	101.250	East Eurasia	Semang
78 Nicobarese	11	1870	7.000	93.750	East Eurasia	Nicobarese
79 Andamanese	5	1860	11.750	93.083	East Eurasia	Andamanese
80 Vedda	6	1860	7.750	81.250	East Eurasia	Indo-Iranian
81 Tanala	11	1925	-22.000	48.000	East Eurasia	Indonesian
82 Negri Sembilan	17	1958	2.583	102.250	East Eurasia	Indonesian
83 Javanese (Miao)	19	1954	-7.700	112.220	Insular Pacific	Indonesian
84 Balinese	18	1958	-8.500	115.333	Insular Pacific	Indonesian
85 Iban	9	1950	2.000	113.000	Insular Pacific	Indonesian
86 Badjau	6	1963	5.000	120.000	Insular Pacific	Indonesian
87 Toradja	11	1910	-2.000	121.000	Insular Pacific	Indonesian
88 Tobelorese	12	1900	2.000	128.000	Insular Pacific	West Papuan
89 Alorese	12	1938	-8.333	124.667	Insular Pacific	Moluccan
90 Tiwi	4	1929	-11.375	131.000	Insular Pacific	Australian
91 Aranda	4	1896	-24.250	133.500	Insular Pacific	Australian
92 Orokaiva	8	1925	-8.500	148.000	Insular Pacific	Central Papuan
93 Kimam	8	1960	-7.500	138.500	Insular Pacific	Central Papuan
94 Kapauku	11	1955	-4.000	136.000	Insular Pacific	Central Papuan
95 Kwoma	11	1960	-4.167	142.667	Insular Pacific	Central Papuan
96 Manus	10	1937	-2.167	147.167	Insular Pacific	Papuan Austronesian
97 New Ireland	10	1930	-2.500	151.000	Insular Pacific	Papuan Austronesian
98 Trobrianders	10	1914	-8.640	151.007	Insular Pacific	Papuan Austronesian
99 Siuai	10	1939	-7.000	155.333	Insular Pacific	Bougainville
100 Tikopia	11	1930	-12.500	168.500	Insular Pacific	Eastern Oceanic
101 Pentecost	9	1953	-16.000	168.000	Insular Pacific	Eastern Oceanic
102 Mbau Fijians	14	1840	-18.000	178.583	Insular Pacific	Eastern Oceanic
103 Ajie	11	1845	-21.333	165.667	Insular Pacific	Papuan Austronesian
104 Maori	10	1820	-35.333	174.167	Insular Pacific	Eastern Oceanic
105 Marquesans	11	1800	-8.917	-140.167	Insular Pacific	Eastern Oceanic
106 Western Samoans	11	1829	-13.750	-172.000	Insular Pacific	Eastern Oceanic
107 Gilbertese	12	1890	3.500	172.333	Insular Pacific	Micronesian
108 Marshallese	12	1900	6.000	168.500	Insular Pacific	Micronesian
109 Trukese	11	1947	7.400	151.667	Insular Pacific	Micronesian
110 Yapese	13	1910	9.500	138.167	Insular Pacific	Micronesian
111 Palauans	12	1947	7.500	134.500	Insular Pacific	Northwest Austronesian

112 Ifugao	12	1910	16.833	121.167	Insular Pacific	Northwest Austronesian
113 Atayal	20	1930	24.333	120.750	East Eurasia	Formosan
114 Chinese	14	1936	31.000	120.083	East Eurasia	Wu
115 Maanchu	18	1915	50.000	125.500	East Eurasia	Tungusic
116 Koreans	20	1947	37.600	126.417	East Eurasia	Korean
117 Japanese	6	1950	34.667	133.667	East Eurasia	Japanese
118 Ainu	8	1880	42.833	143.000	East Eurasia	Ainu
119 Gilyak	7	1890	54.000	142.500	East Eurasia	Gilyak
120 Yukaghir	7	1850	64.750	153.500	East Eurasia	Yukaghir
121 Chukchee	11	1900	66.500	180.000	Insular Pacific	Chuckhee-Kamchatcha
122 Ingalik	8	1885	62.500	-159.500	North America	Northern Athabaskan
123 Aleut	9	1800	55.250	-164.000	North America	Aleutian
124 Copper Eskimo	5	1915	68.000	-112.500	North America	Eskimoan
125 Montagnais	6	1910	50.000	-74.000	North America	Algonquian
126 Micmac	7	1650	46.000	-63.000	North America	Algonquian
127 Saulteaux (Ojibwa)	8	1930	52.000	-95.500	North America	Algonquian
128 Slave	6	1940	62.000	-122.000	North America	Athabaskan
129 Kaska (Nahane)	3	1900	60.000	-131.000	North America	Athabaskan
130 Eyak	7	1890	60.500	-145.000	North America	Eyak
131 Haida	10	1875	54.000	-132.500	North America	Haida
132 Bellacoola	9	1880	52.333	-126.500	North America	Salishan
133 Twana	9	1860	47.433	-123.250	North America	Salishan
134 Yurok	8	1850	41.500	-124.000	North America	Ritwan
135 Pomo	9	1850	39.000	-123.000	North America	Pomo
136 Yokuts	9	1850	35.000	-119.500	North America	Yokut
137 Paiute (Northern)	5	1870	43.500	-119.000	North America	Shoshonean
138 Klamath	6	1860	42.625	-121.667	North America	Sahaptin
139 Kutenai	7	1890	49.000	-116.667	North America	Wakashan
140 Gros Ventre	7	1880	48.000	-108.000	North America	Algonquian
141 Hidatsa	10	1836	47.000	-101.000	North America	Siouan
142 Pawnee	10	1867	42.000	-100.000	North America	Caddoan
143 Omaha (Dhegiha)	11	1860	41.433	-96.500	North America	Siouan
144 Huron	12	1634	44.500	-79.000	North America	Iroquian
145 Creek	13	1800	32.933	-86.000	North America	Natchez-Muskogean
146 Natchez	13	1718	31.500	-91.417	North America	Natchez-Muskogean
147 Comanche	7	1870	34.000	-101.500	North America	Shoshonean
148 Chiricahua	6	1870	32.000	-109.500	North America	Southern Athabaskan
149 Zuni	14	1880	35.667	-108.750	North America	Zuni
150 Havasupai	8	1918	35.833	-112.167	North America	Yuman
151 Papago	11	1910	32.000	-112.000	North America	Sonoran
152 Huichol	11	1890	22.000	-105.000	North America	Sonoran
153 Aztec	16	1520	19.000	-99.167	North America	Aztecan
154 Populuca	11	1940	18.250	-94.833	North America	Oto-Manguean
155 Quiche	14	1930	15.000	-91.000	South America	Mayan
156 Miskito (Mosquito)	10	1921	15.000	-83.000	South America	Misumalpan
157 Bribi (Talamanca)	9	1917	9.000	-83.250	South America	Western Chibchan
158 Cuna	13	1927	9.250	-78.500	South America	Western Chibchan
159 Goajiro	8	1947	11.917	-71.750	South America	Arawakan
160 Haitians	17	1935	18.833	-72.167	South America	Romance
161 Callinago	10	1650	15.500	-60.500	South America	Cariban

162 Warrau (Warao)	5	1935	9.078	-62.000	South America	Warrauan
163 Yanomamo	7	1965	2.417	-65.000	South America	Yanomaman
164 Carib	7	1932	7.417	-60.167	South America	Cariban
165 Saramacca	11	1928	3.500	-55.750	South America	Romance
166 Mundurucu	8	1850	-6.500	-56.500	South America	Tupi-Guarani
167 Cubeo (Tucano)	8	1939	1.250	-70.500	South America	Tucanoan
168 Cayapa	9	1908	1.000	-79.000	South America	Paezan
169 Jivaro	6	1920	-3.000	-78.000	South America	Jivaroan
170 Amahuaca	6	1960	-10.333	-72.250	South America	Panoan
171 lnca	14	1530	-13.500	-72.000	South America	Quechuan
172 Aymara	13	1940	-16.000	-65.750	South America	Quechuan
173 Siriono	4	1942	-14.500	-63.500	South America	Tupi-Guarani
174 Nambicuara	6	1940	-13.000	-58.750	South America	Ge
175 Trumai	7	1938	-11.833	-53.667	South America	Timote
176 Timbira	9	1915	-6.500	-46.000	South America	Ge
177 Tupinamba	9	1550	-22.792	-44.500	South America	Tupi-Guarani
178 Botocudo	2	1884	-19.000	-42.500	South America	Botocudo
179 Shavante	5	1958	-13.500	-51.500	South America	Ge
180 Aweikoma	2	1932	-28.000	-50.000	South America	Ge
181 Cayua (Guarani)	6	1890	-23.500	-55.000	South America	Tupi-Guarani
182 Lengua	7	1889	-23.000	-58.500	South America	Mascoian
183 Abipon	8	1750	-28.000	-59.500	South America	Guaycuran
184 Mapuche	12	1950	-38.500	-72.583	South America	Araucanian
185 Tehuelche	6	1870	-40.500	-68.000	South America	Tehuelchan
186 Yahgan	2	1865	-55.500	-69.500	South America	Yaghan

9. APPENDIX 2. ADOLESCENT INTIATION CEREMONIES (SCHLEGEL AND BARRY 1979)

529. Occurrence: Boys
530. Occurrence: Girls

	Boys	Girls
. = Missing data	4	3
1 = Absent for both boys and girls	80	81
2 = Absent for specified sex only	39	17
3 = Present	63	85

531. Time: Boys
532. Time: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	120	100
2 = Before genital maturation	13	9
3 = At first signs of genital maturation	18	11
4 = At genital maturation	6	57
5 = Within one year after genital maturation	17	5
6 = Later (up to 18 years)	8	1

533. Number of Concurrent Initiates: Boys
534. Number of Concurrent Initiates: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = Single	29	73
3 = Small group	7	6
4 = Large group	27	5

535. Duration of Ceremony: Boys
536. Duration of Ceremony: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = Short	28	36
3 = Medium	7	21
4 = Long	28	27

537. Number of Participants: Boys
538. Number of Participants: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	121	99
2 = Immediate family	7	40
3 = Local group	25	29
4 = Large group	29	15

539. Sexes of Participants: Boys
540. Sexes of Participants: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = Both sexes	12	11
3 = Partially limited to same sex as initiates	17	28
4 = Exclusively same sex as initiates	34	45

541. Primary Physical Components: Boys
542. Primary Physical Components: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = None	6	11
3 = Manipulations or activities	17	45
4 = Pain other than genital operation	20	21
5 = Genital operation	13	7
6 = Genital operation and other pain	7	0

543. Secondary Physical Components: Boys
544. Secondary Physical Components: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = Neither manipulations nor activities	15	20
3 = Activities	14	10
4 = Manipulation	9	26
5 = Both manipulations and activities	25	28

545. Primary Cognitive or Performance Components: Boys
546. Primary Cognitive or Performance Components: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = Symbolic only	20	15
3 = Learning skills, sharing secrets, or other	3	3
4 = Observing taboos	8	1
5 = Seclusion	7	9
6 = Both seclusion and observing taboos	18	54
7 = Fear	7	2

547. Secondary Cognitive or Performance Components: Boys
548. Secondary Cognitive or Performance Components: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	118	100
2 = Neither learning skills nor sharing secrets	43	60
3 = Sharing secrets	8	2
4 = Learning skills	4	11
5 = Both learning skills and sharing secrets	9	10

549. Primary Emic Interpretations: Boys
550. Primary Emic Interpretations: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	119	99
2 = None	4	5
3 = Status marker, physical change, or behavior change	41	75
4 = Spiritual change	11	2
5 = Death-rebirth	7	2

551. Secondary Emic Interpretations: Boys
552. Secondary Emic Interpretations: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	118	100
2 = No status marker	8	8
3 = General status marker	17	25
4 = Status marker for adolescence or youth	14	12
5 = Status marker for full adulthood	25	38

553. Tertiary Emic Interpretations: Boys
554. Tertiary Emic Interpretations: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	118	100
2 = Neither physical nor behavior change	31	48
3 = Behavior change	10	12
4 = Physical change	12	16
5 = Both physical and behavior change	11	7

555. Primary Social Consequences: Boys
556. Primary Social Consequences: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	118	100
2 = None	19	32
3 = Familial integration, familial independence, or other	14	20
4 = Heterosexual intercourse	8	25
5 = Same-sex bonding	17	3
6 = Both same-sex bonding and heterosexual intercourse	6	3

557. Secondary Social Consequences: Boys
558. Secondary Social Consequences: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	118	100
2 = None	36	57
3 = Other	6	8
4 = Familial independence	13	9
5 = Familial integration	9	9

559. Principal Focus: Boys
560. Principal Focus: Girls

	Boys	Girls
. = Missing data	4	3
0 = Absent	120	111
2 = Fertility	11	34
3 = Sexuality	10	18
4 = Valor	6	1
5 = Wisdom	7	1
6 = Responsibility	26	23
7 = Other	2	7

.