

Kent Academic Repository

Full text document (pdf)

Citation for published version

Ember, Carol and Fischer, Michael D. (2017) Using eHRAF World Cultures with other cross-cultural samples. In: White, Douglas, ed. Wiley Cross-cultural Research Companion. Wiley-Blackwell, NY USA. ISBN 9781118499399. (Submitted)

DOI

Link to record in KAR

<http://kar.kent.ac.uk/58166/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

For Wiley Cross-Cultural Companion

Using eHRAF World Cultures with other cross-cultural samples

Carol R. Ember

Human Relations Area Files at Yale University

Michael Fischer

University of Kent

The Human Relations Area Files (HRAF) was founded in 1949 as a membership consortium with the goal of facilitating the comparative study of human behavior. In addition to carrying out cross-cultural research at the HRAF headquarters, a principal program was to physically bring together ethnographic information about the cultures of the world so that researchers in other institutions and in fields outside of anthropology could efficiently learn about human cultures and then formulate and test hypotheses across human cultures in all their variety. The precursor to the HRAF collection of ethnography—the “Cross-Cultural Survey”—was developed at the Institute of Human Relations at Yale, an interdisciplinary research center. (For an overview of HRAF’s history see Ember (2012)). The original collection of paper files in file cabinets was commonly referred to as the “HRAF Files.” We more formally refer to the whole collection (in paper, microfiche, and now mostly online), as the HRAF Collection of Ethnography.

What was unique about the enterprise was not just putting ethnographic texts and other material together in one place; it was the decision to finely index entire ethnographic works at the paragraph-level to support finding related information from multiple ethnographies. This helps researchers code variables relatively quickly, and provides avenues for other approaches to comparative ethnographic analysis, quantitative and qualitative. Subject-indexing was based on HRAF’s subject classification system, the Outline of Cultural Materials (OCM) (Murdock, 2008). The OCM is a thesaurus with over 750 scope notes, broader and narrower terms, natural language pointers to the subject categories, and related terms. HRAF was founded before the computer age and even before copying machines. Graduate students had to painstakingly re-type individual paragraphs and make 6 copies at a time with onion-skin paper. The HRAF Files were a pre-computer technological innovation to facilitate rapid retrieval of similar subjects across cultures (Ember, 2012:619). In the paper version, pages were duplicated as many times as there were subjects on the page and pages were then ordered and then filed in a cabinet drawer by subject. So, for example, if you wanted to read about “Techniques of Socialization” you could go to the file drawer of a particular culture, go to category 861 (a shorthand for first subject in the 86, or socialization subject-categories), and read

all the pages by all the authors who wrote about the same subject. At the top of every page was a “header” containing some metadata, including the field date, the time coverage, and the date of publication. (HRAF has always tried to include information about more than one time period and more than one community or region within the same culture.) After there were copy machines, trained indexers (usually anthropologists) wrote these shorthand codes in the margins next to the original paragraphs. These numeral “codes” are often misunderstood to be “coded” data. However, they are not codes in the usual sense; these are best thought of as pointers to relevant subject matter.

Not surprisingly, the transition to digital format (now eHRAF World Cultures) has eliminated the necessity for file cabinets and the need to physically separate pages into the same subject. And digital searching now also allows searching by keywords. But the main purpose of eHRAF remains the same—to facilitate coding of variables to test hypotheses, although as the entire body of underlying texts is now accessible, new opportunities for ethnographic analysis will emerge that leverage the codes in association with the texts.

While digital searching now allows researchers to use keywords for searching, this has unfortunately led some researchers to prefer keyword searching to using the OCM categories. While word searching is arguably better for some topics, particularly those where there is limited variation in word usage (“irrigat*” for irrigation or irrigate vs. “plow* plough*”), keywords result in very poor searches for concepts that are often expressed by a large number of terms and idioms. Putting in the root “punish*” combined with the root “child*” will find instances of paragraphs where only those words are used, but will miss a very large number of described instances when ethnographers use other phraseology such as “parents rarely reprimand their children” or “a grandmother will scare a child with stories about how an owl will carry them off if they are naughty.” In contrast, a succinct OCM category such as “Techniques of Socialization” (OCM 861) usually will contain appropriate information regardless of the particular wording used by the ethnographer. While HRAF is currently experimenting with methods for improved keyword and topic based searching, even once implemented using OCM codes will likely greatly improve search results.

The HRAF Collection of Ethnography provided a major boost to a researcher’s ability to conduct cross-cultural research in a reasonable time and such studies grew fairly dramatically in the 1960s and 1970s. This can be illustrated by counting the number of published cross-cultural studies per decade in HRAF’s database of cross-cultural research findings, *Explaining Human Culture* (<http://hraf.yale.edu/ehc/>).¹ The number of cross-cultural studies started to climb dramatically in the 1960s, peaked in the 1970s, dropped in the 1990s, and appears to be back on the increase in the 21st century. Obviously there are other factors at work than the presence of the HRAF Collection of Ethnography.

Most notably there was a pronounced lack of interest in cross-cultural research coinciding with the post-modern movement, which not only rejected comparative research, but also ethnography as a source of data beyond the author's unique experience.) Second, the development of the five C's project for coding the Standard Cross-Cultural Sample in Pittsburgh in the early 1970s probably contributed to an increase in cross-cultural research in the next two decades. Table 1 shows that some form of the "HRAF Files" has fairly consistently been employed in about 1/3-1/4 of the cross-cultural studies in each decade, while the SCCS became more important in the two decades since its appearance. This is probably due to the fact that the SCCS sample has a very large number of already coded variables.

[Table 1 near here]

Samples and Comparative Research

The HRAF Collection of Ethnography supports a broad range of comparative research, much of it qualitative where students and scholars use the eHRAF application to identify and collate relevant ethnographic information from a number of societies on a topic of interest, such as child socialization, which they then read and compare between societies. Most of this is directed to user selected samples, say societies in south eastern Africa, two specific societies that are adjacent geographically, or even early, mid and late ethnographic sources for a single society. The eHRAF application has a number of features that make selection at the society level easy, and thus fairly easy to construct any arbitrary sample of societies. The sample, in this sense, is constrained to meet the requirement and goals of the research in the judgement of the researcher.

One of the issues with this approach is that while very interesting results can be achieved, it is difficult to set these out in the context of cultures and societies more generally. That is, the results may be interesting and valid, but there is little support for generalization to other societies, developing a sense of pervasiveness or identifying the dynamics of variation without recourse to a more systematic approach.

One common approach to a more quantitative means of assessing comparative research, often referred to as cross-cultural research, is to use a pre-defined sample, often intended to be constructed of societies that are relatively independent of each other, represent different technological or economic levels, are dispersed geographically, or some combination of these. The idea underlying the construction of a sample is that conventional statistics can be used to assess the strength and significance of any correlations or association found between different properties of the societies, and general statements can be made within the limits of the sample's construction. These properties are usually represented as coded variables, where 2 to N possible values are construc-

ted and assigned to each variable (see section **Coding Variables from the HRAF Collection of Ethnography** following). For example, HAS TIES TO STATE might have values yes, no and missing, or PROPORTION OF ECONOMY BASED ON AGRICULTURE which might range from 0 to 100%. The capacity for generalization is limited to the basis upon which the sample was constructed, which if standard statistics are to be used, must conform to assumptions relating to independence and the distribution of specific variable values.

Most researchers use this approach as the construction, interpretation and design of appropriate statistics is difficult, while the approach presented here is relatively well supported through a number of existing samples, and important advances have been made to improve conformance to the underlying assumptions. (Editor: [link to Doug's project in this volume](#)).

How the HRAF Collection of Ethnography Compares with Other Samples

The first major difference between the HRAF Collection of Ethnography (HRAF Collection for short) and other cross-cultural samples is that most of the commonly used cross-cultural samples have one time and a place (or an ethnographer) focus. The HRAF Collection does not typically have one focus per culture; in most cases the HRAF Collection purposely includes a range of ethnographies and range of communities for each society. Unless a hypothesis specifies a diachronic relationship, it is critical to match time foci (and preferably place foci) for all the variables studied (Divale, 1975, Ember et al., 1991). The reason is that if the time and place foci are not controlled, measurement error is likely to increase and correlations are likely to decline in magnitude and representiveness. However, since the HRAF collection includes multiple textual sources and metadata relating to each society, it becomes possible to create refined samples and address complex research problems depending on both synchronic and diachronic considerations.

A second major difference is that while the HRAF Collection may have been started intending to be a representative sample of the societies of the world, it, in contrast to other samples, is not currently recommended (Human Relations Area Files) as a sampling universe in its entirety when, for example, desired results depend on case independence or randomized geographical distribution. This is because the HRAF Collection has over time grown in some opportunistic ways (such as including immigrant cultures in North America which will not be independent of the origin culture) and selection of cases based in part on the depth of documentation available. This adds value for many research problems, but requires greater sampling focus for problems where quantitative comparative analysis requires independence of cases or geographical distribution, or there is a need to meet other assumptions, for example, cases being normally

distributed across some dimension. See the next section for sampling suggestions.

Most of the other samples claim to be either fairly exhaustive of some sampling universe or representative of the world's societies in some way. These samples include: 1) the second edition of the Standard Ethnographic Sample (Naroll and Sipes, 1973) which claims to include all the "primitive" societies that met Naroll's data quality criteria; 2) the Ethnographic Atlas (Murdock, 1962–1971) which claims, with some caveats, to be fairly close to a sampling universe; 3) the Standard Cross-Cultural Sample (Murdock and White, 1969) which claims to have a representative well-described society from each of 186 culture areas; 4) the Atlas of World Cultures (Murdock, 1981) which claims to include a majority of the most fully described cultures; and 5) the Probability Sample Files (Lagacé, 1979, Naroll, 1967), a subset of the HRAF Collection that has a randomly selected society from each of 60 culture areas. (Ember et al., 1992)

A third difference is that the HRAF Collection as a whole does not employ any methods of stratification by culture area or geographic area as did the Atlas of World Cultures and the SCCS.

A fourth difference is that the HRAF Collection is not a fixed sample, but is annually growing and updated with new societies and additional ethnographic sources

Therefore, the HRAF Collection is not intended to be a sample as such, as the criteria and assumptions used for selecting the cases are unlikely to match a priori the characteristics needed for a given analysis. However, cases can be matched against existing samples, or new samples with appropriate characteristics drawn from the HRAF Collection. The HRAF Collection is best considered a resource that can be used to support analyses based on samples.

Sampling Considerations Using the HRAF Collection of Ethnography

Since the HRAF Collection of Ethnography was not developed as a sample for testing, but rather as a source to inform comparative investigations, we do not recommend its use in its entirety as an appropriate sample for scientific testing of hypotheses that presume a representative sample. However, we do recommend one of two strategies for sampling from the cultures documented in the HRAF Collection. The first is to use the 60-culture subset of the HRAF Collection—the Probability Sample Files (PSF), perhaps supplemented by HRAF's Simple Random Sample (societies chosen randomly from the over 2,000 societies listed in the Ember et al.'s concordance (1992)). The second strategy is to use another sample that claims to be representative and wherever possible use the cases in eHRAF World Cultures (eWC—the online version of the collection) for coding, and, if appropriate to the problem, substitute cultures from eWC for

cultures in the sample, although this should be done with care depending on the issues under research. As of July 2016, the sample with the highest overlap is the SCCS sample (over 75 percent of the societies are included) and HRAF plans to include all the societies in the SCCS by the end of June in 2020.

Which sample to use may depend on the variables that are needed for research. If the intention is to code all new variables, the 60-culture PSF sample, now selectable in the eHRAF World Cultures application, may be an ideal size. If the researcher wants to use some variables from another sample, the choice depends upon the variables that are needed and available. In the next section, we discuss how eHRAF can be used for coding and how time and place foci can be matched to other samples.

Using the HRAF Collections with Coded Variables from Other Samples

As mentioned above, most of the published cross-cultural samples contain ethnographic references, but the HRAF Collection is unique in actually containing ethnographic material. It does not presently retrieve pre-coded variables for any sample. The value for cross cultural research using external samples is support for investigating further the pre-coded variables from other samples, or for providing an ethnographic basis for coding additional variables. The HRAF Collection also differs in having multiple time and place foci for each society. But since cultures change, almost all coded variables have a time and a place focus. So how does one use the HRAF Collection of Ethnography and at the same time focus on the same time and place as the sample whose variables you want to use?

Meaning of a Match Between Samples

The matches between the HRAF Collection and 7 other cross-cultural samples were part of a larger concordance database project conducted primarily in 1991 (Ember, et al., 1992). The other samples include: the World Ethnographic Sample (Murdock, 1957), the entire Ethnographic Atlas ((Murdock, 1962-1971), the Probability Sample Files (also called the HRAF Quality Control Sample), the Standard Cross-Cultural Sample (Murdock and White, 1969), the Standard Ethnographic Sample, second edition (Naroll and Sipes, 1973), and the Atlas of World Cultures (Murdock, 1981). By "concordance" Ember et al. (1992: 3) meant "that as far as we can determine the data from both samples pertain to the same time and place foci." The concordance database does not contain one record per society; rather, a record is a specific unique time and place focus (defined as a unique 10-year time period and a different place); some cultures have many records if there are many described foci. (Note that a focus does not have to be specific; if an ethnographer describes a society generally it is considered a different focus from a named community or regional focus; similarly, an unspecific time focus is also considered a different time focus.) Each

record contains identifying fields pertaining to each of the cross-cultural samples and if it is a match than it will be considered present in that sample. Since the cultures in the HRAF Collection almost always contain multiple foci, the concordance lists the particular documents that pertain to each of the foci. In the next two sections we describe how to find matches between eHRAF World Cultures and the SCCS and EA samples.

SCCS matches

By design the HRAF Collection does not usually focus on only one time and one place, so a researcher needs to take steps to match foci when coding or using variables from others. HRAF has done these matches for the matches between eHRAF World Cultures and the SCCS. Note, that although HRAF intends to include all the SCCS cases within the next four years (July 2020), not all the cases are yet included. In addition, a few cultures in both samples do not as yet have matching documents (e.g., Basques). There are two ways that matches are provided. The first is in a public table that will be updated as new eHRAF cases are added on HRAF's home page (<http://hraf.yale.edu/resources/reference/sccs-cases-in-ehraf/>). See Figure 1 for a portion of the results on SCCS matches. The current SCCS cases in eHRAF are shown with permalinks to the matching documents in eHRAF. The second method is that after performing a search in eHRAF you can "Narrow or Filter" the results to show you only those cultures for which there are at least some matching documents to the SCCS sample cases (see Figure 2). Then, in the Document and Paragraph Results View for an SCCS case, matching SCCS documents will be footnoted with "S1," "S2," and "S3." An "S1" is considered a direct match. An "S2" is considered a partial match, usually because the document covers multiple time periods or multiple cultures. An "S3" footnote is occasionally given when the time or place is somewhat ambiguous and we urge some caution in using the source. No footnote beside the document title may mean one of two things: 1) that while some document in the culture is a match for the SCCS, the particular document listed is not considered a time and place match; or 2) none of the matching documents have information on the subject you are looking for. We still include non-matching results in a search in order that researchers have access to additional data that may be relevant to them despite not being an exact SCCS match.

[Figures 1 and 2 near here]

If you are still using the HRAF Collection in paper or microfiche, Table 2 shows you the matches between the societies not yet in eHRAF World Cultures (or with no matching documents) and the SCCS sample.

[Table 2 near here]

Ethnographic Atlas matches

[Note to editor: the changes described here are not yet implemented, but will be by the time the chapter is published]

The Ethnographic Atlas (EA) also has a time and place focus for each culture. HRAF will provide two methods to finding matching documents in eHRAF World Cultures. The first is in the form of a table on our public home page; the second is a way to narrow or filter to EA cases after performing a search. The procedure will be the same as shown in Figure 1 except that there will be a check box for the Ethnographic Atlas.² The EA is a very large sample with over 1200 societies and therefore, in contrast to the SCCS, HRAF has no current plans to include the whole EA sample in eHRAF World Cultures. Having the matches will be particularly useful if you want to use some coded data from the EA. (The online database D-PLACE-- <https://d-place.org/home>-- which includes codes from the EA--has links to cases included in eHRAF World Cultures. See <http://hraf.yale.edu/hraf-collaboration-with-d-place/> for a discussion of the linkages between eHRAF and D-PLACE.)

Note that the EA and the SCCS foci are not always the same even if both samples contain the same cases. For instance the focal time for the Ethnographic Atlas is the 1920s based on the fieldwork of Margaret Mead and Peter H. Buck in American Samoa, whereas the SCCS focal time is in 1829 based on the work of George Turner and John Stair on Western Upolu island in Western Samoa. (Ember, 2007:397).

Coding Variables from the HRAF Collection of Ethnography

Coding your own variables may be time-consuming, but the process is very rewarding. First, cultures around the world have an amazing variety of cultural practices, beliefs, and values and even though cross-cultural researchers accept that societies can be compared and scaled on certain dimensions, or share some feature in the general sense, reading about how particular people actually do things is quite simply fascinating. Second, from a methodological point of view, without having coded at least a portion of the societies in your sample yourself, you don't have a sound basis for deciding if a particular code (yours or from others') is what you want or need. Third, if one or more of the hypotheses you are testing is not supported, lack of familiarity with any cases makes it difficult to come up with alternative theories. On a personal note, which is why the first author is a strong advocate for doing at least some of your own coding, the Embers' (Carol and Melvin) strongest cross-cultural findings have come from research in which at least some of the variables for the study were specially coded for the research project. A case in point is the Ember and Ember (1992) study on the relationship between resource problems and frequent warfare. The Embers were trying to test theories about population pressure on resources leading to more warfare. They first looked for possible measures that

were already coded, but although there were some scales that might possibly have suited, such as Murdock and Morrow's (1970) code for preservation and storage of food that appeared to describe adequacy of food storage, the Embers were not sure that the measure was close enough to what was needed, so they decided to code three new measures indicating that populations had overstepped their carrying capacity. Using a 25-year time period around the ethnographic present, three variables were coded: "chronic food scarcity," "frequency of famine," and "natural hazards that seriously destroyed food supplies." It turned out that not only were natural hazards an extraordinarily strong predictor of more warfare in non-state societies, but the Embers' measures did not correlate well with the Murdock and Morrow measures that they originally considered. This confirmed their initial judgment that the food storage variable was not an appropriate measure for the theoretical constructs they were trying to measure.

Designing appropriate variables to measure concepts from ethnography should be a very iterative process because it is not clear without a good deal of trial and error that coding directions or questions are clear enough to achieve reliability. But it may also happen that the information you may want is not to be found in a sufficient number of cases. For instance, you might want to know the percentage of households that are extended family households in a community, but after reading a number of cases, you may realize that few ethnographers give you the quantitative information you would like. At this point you have a choice (Ember and Ember 2009, 46–50)—you can abandon your quest or you may choose to redesign your coding scale based on qualitative assessments such as "the typical household is a large extended family." Note that making this decision does not preclude you from measuring more precisely wherever possible; in fact, Ember and Ember (2009:52) recommend that you aim for a precise measure and a less precise measure. Another alternative is to use a data quality score that reflects the type of information provided by the ethnographer and you can reanalyze the results with and without the scores based on lower quality data.

Two other important steps are: 1) choosing a time and a place focus; and 2) providing detailed instructions as to how to find the information you are seeking. If you are coding from the HRAF Collection in conjunction with variables from another sample, the time and place focus choice is clear—as discussed above, you need to use the focus specified in the other sample. But if you are coding all the variables from the HRAF Collection, you should choose a focus based on either theoretical grounds (such as wanting a focus with minimal effects of colonialism) and/or choosing an ethnographer who best describes the topic matter you are interested in (Ember and Ember, 2009:76–78).

eHRAF World Cultures allows multiple search strategies (by paragraphs containing particular OCM subjects and/or keyword) and part of designing a coding

scheme is to specify the procedures you want followed to find information. As discussed above, word searching is often unduly narrowing, especially when ethnographers can use alternative words. That is why we usually recommend finding one or more OCM subjects that are likely to include the appropriate information. One of the most important things to keep in mind is that coding rules also have to specify how to find that a trait or custom is absent. Suppose you are trying to determine the degree to which a society has extended families. At first glance, it might seem appropriate to have the coders look at the category “Extended Families” (OCM 596). But what if there is no information in that category? This may not mean that extended families are absent because the ethnographer might not have attended to family structure. It is prudent to instruct coders to look at other categories where any kind of family might be described. “Household” (OCM 592) might have been a better choice since it is about the general composition of the household or family compound. If there is little or no information there, the coder would be able to conclude that information is lacking on this subject.

The other problem with word searching is that it usually only finds positive instances, not negative instances. Ethnographers are likely to tell you if a community irrigates their fields when they observe it, but they usually do not remark on all the things that people do not do.

So coding instructions for extended family might say “Use Advanced Search with subject category ‘Household (OCM 592)’ focusing on the time and place (or ethnographer name) specified on the code sheet for the sampled society.” If the category contains sufficient information to code the degree to which extended families are present, code the case. If there is not sufficient information, look at other family subject categories, “Nuclear Families,” “Polygamy,” and “Extended Families.”

Strategies for Incomplete or Custom Samples

Unless you use the HRAF 60 society probability sample, the HRAF Collection will only assist you with documentation on a portion of your sample. There are a variety of ways you can address this, including:

1. Finding other sources. Most samples include a bibliography for each society in the sample. Other bibliographic sources for a society with the same time and space focus may be used as a basis for coding.
2. Substitution. Some samples may include criteria for each society that facilitate substituting another society. For example, in the Standard Cross-Cultural Sample each society was the best documented society in each of 186 cultural areas at a time when independence from other cultures was near a peak. Substituting another well documented society from the same cultural

region with an appropriate time focus should not perturb results in a significant manner.

3. Working with the partial sample. An assessment of impact of missing cases should be made relative to the research problem. For example, if you are comparing sub-Saharan to Circum-Mediterranean societies, missing Asian societies will not impact the result. There is no one way to make this assessment. For example, you can compare the results using related pre-coded variables on the entire sample and the part of the sample that overlaps eHRAF. The extent to which these are similar can be used to guide your interpretation of results from the partial sample. If you are looking only at a sub-sample of hunter/gatherer societies you need only assess the impact of missing hunter/gatherer societies.

The more pressing issues are coding new entries, auditing or investigating the basis for existing coding, and to evaluate what can safely be concluded as a result of the partial state of the target sample.

Where the sample society is present, the HRAF Collection supports coding as in the previous section. If missing entries are to be completed, these must be coded corresponding to other sources drawn from the documentation for that sample, unless you are substituting other societies from eHRAF. If you are working with a partial sample comprised of cases in eHRAF, to make the presentation of results comparable to studies using the full sample you can code a missing value for each variable used for each of the societies not present. How a missing value is entered, and indeed whether you need to enter these, depends on the software you intend to use to analyze the results.

As of April 2016, there were 53 cases missing from the SCCS sample, which are reasonably well dispersed throughout the sample (see Figure 3 where a square represents a missing SCCS case.³ We expect to be on complete parity by mid-2020. The two main culture areas where serious distortion will likely occur are Mesoamerica (missing 4 consecutive SCCS cases, 152–155), and to some extent, Melanesia (missing cases 101–103), where there are relatively large numbers of missing societies given the small number of cultures in these areas. Otherwise the eHRAF SCCS sub-sample is representative of the full sample for most variables.

[Figure 3 near here]

To evaluate whether or not using the partial eHRAF SCCS sample does undue harm in testing hypotheses, we have done a couple of comparisons of statistical results using the whole SCCS sample compared with the subset of those included in eHRAF (cases shown in Figure 3). Table 3 shows these comparisons. These can be accessed from <http://lucy.kent.ac.uk/Ethnoatlas/SCCS> and <http://lucy.kent.ac.uk/Ethnoatlas/hrafSCCS> respectively, reflecting updated HRAF SCCS coverage. [Note to editor. URLs will change before publication.] The

first example in Table 3 is the cross-tabulation between hunting as a proportion of the economy and gathering as a proportion of the economy. Most SCCS variables have 10 or fewer possible values. In this case each value represents an interval of the proportion of the economy due to either hunting or gathering. In the examples, these values have been merged into three categories, corresponding to low/medium/high proportions. Missing values have been excluded, although there were none for the full sample, there were the values corresponding to the 53 missing cases for the HRAF sample.

[Table 3 near here]

Note that the observed data is distributed relatively proportionally between the SCCS and the HRAF SCCS cases, remaining fairly close to the 10/7 ratio between the two samples. The result in both supports the hypothesis that as dependence on gathering increases there is a corresponding increase in hunting. The similarity in significance of each distribution is confirmed by examining the corresponding contributions to χ^2 for each cell. The most sensitive cells are in row 3, where even in the original two of the cell frequencies are marginal since small differences in data frequencies can result in large changes in χ^2 . On balance these average out, and the result is a lower total χ^2 for the HRAF SCCS, but still well within the .05 confidence interval commonly used in the social sciences. The second example in Table 3 is the cross-tabulation between the form of betrothal with the proportion dependence on agriculture as a proportion of economy. We find similar results to the first example, even given we are working with a subsample in both results in the second example. In the SCCS, form of betrothal has seven possible values, and in this case only two of these are being examined. This excludes nearly half of the 186 cases, leaving 95 cases for consideration. Distributing these values over low and high reliance on agriculture, the observed data from both the SCCS and HRAF SCCS supports a hypothesis that brideprice or bridewealth is associated with higher dependence on agriculture, and dowry associated with lower dependence on agriculture. Although the cell contributions to χ^2 are relatively smaller for HRAF SCCS, these are still well within the .05 confidence interval. In both examples the χ^2 values tend to be overall lower for the HRAF SCCS, although it is possible in individual cells for it to be higher, as in the first example in Table 3 in the case of high levels of gathering with mid-levels of hunting. But overall the strong tendency will be for the HRAF SCCS sample to understate χ^2 relative to the full sample because all the cases in the HRAF SCCS are in the full SCCS, while the missing cases will make cell frequencies lower for the HRAF SCCS overall, which will usually result in a lower overall χ^2 [Note to editor—the symbol χ^2 in this paragraph is the Greek letter chi followed by a superscript 2 standing for chi-square.].

[Table 3 about here – should appear as opposite page or previous page]

There are issues that arise in evaluating the results of quantitative cross-cultural analysis, even from any full corresponding sample, such as the HRAF

Probability Sample, but particularly in whatever partial sample can be drawn from eHRAF for samples such as the SCCS. One issue of concern to some cross-cultural researchers is possible autocorrelation between societies, particularly that arising from cultural diffusion due to geographic proximity, conquests and adventures. The significance of autocorrelation in the results of cross-cultural analysis has been a subject of debate for some time,⁴ particularly arriving at a quantitative measurement. The case for both its significance, how to quantitatively measure it, and how to limit the influence of autocorrelation can be found in several entries in this volume (to editor: [REFs from this work HERE](#)).

Even without considering autocorrelation the assumption of conformance to a random distribution can be a problem for determining the significance of a result. Certainly using a partial sample of a sample drawn on the assumption of random assignment is problematic in this respect, even if we have some confidence in the subsample itself, simply because we can not safely evaluate significance.

However, within the limits of conventional analysis, the HRAF SCCS sample should be useful for a wide range of problems where it corresponds well to the full SCCS. Other methods proposed in the volume can only improve these results.

Discussion and Conclusions

The HRAF Collection of Ethnography (in paper, fiche, and online) was designed to facilitate cross-cultural coding, but designing and coding new variables is much more time-consuming than using coded data from other researchers. And it is therefore very tempting to use precoded data. But often compromises are made if the variables are not quite measuring the concept that is desired. The important point is that the more the measure departs from the theoretical concept, the more measurement error is likely. And measurement error almost always makes it difficult to find a “true” relationship (Ember and Ember, 2009:54–55). But even if the construct measured is what you want, without examining at least some of the coding it is difficult to be sure that the previous researcher followed the methodology you would have wanted. For example, was there a clear time and place focus? How much guessing was allowed? When the Embers were retesting John Whiting’s theories about male initiation for the SCCS, they first decided to use codes from Barry and Paxson (1971) to examine the possible effects of mother-child sleeping and the long post-partum sex taboo. But the correlations were much lower than those obtained by Whiting, so Carol Ember decided to recode those variables from the ethnographic literature. In doing so, she paid close attention to the time and place focus and also declined to code many of the previously coded cases because of ambiguous information. Although the sample size was reduced, the correlations improved substantially using the new coding and closely matched those

originally found by Whiting (see the discussion in Ember and Ember (2010)). This example illustrates the importance of at least recoding a portion of the cases done by another researcher to see if you are comfortable with their codes for your project.

Recognizing that the HRAF Collection of Ethnography does not comprise a sampling frame for hypothesis-testing using inferential statistics, HRAF created a 60-culture sample called the Probability Sample Files with one randomly selected culture that met data quality standards from each of 60 culture areas. As mentioned above, this sample can be supplemented by the Simple Random Sample (SRS) identified in eHRAF World Cultures. As we have demonstrated, even the present subsample of the SCCS in eHRAF World Cultures can be usefully applied to a range of problems. However, since the SCCS is now the most commonly used sample, the HRAF Board approved a plan to incorporate all the SCCS cases by the end of June 2020 and, as we explained above, we have now implemented an easy way for researchers who want to use the SCCS and also code variables from eHRAF World Cultures to find the documents that match the time and place foci in the SCCS. A similar feature will be added to match to the Ethnographic Atlas shortly.

In conjunction with HRAF's new database, *Explaining Human Culture* (<http://hraf.yale.edu/ehc>), which summarizes the results of previous cross-cultural research, it is our hope that these coordinated resources will facilitate and encourage cross-cultural research to a greater degree.

NOTES

Note that this database, released in June 2016 with over 800 studies, contains a substantial number of cross-cultural studies that HRAF has located, but it does not claim to be complete. The cross-cultural studies in the database involve at least one descriptive or relational hypothesis on at least 10 or more primarily anthropological cultures. We included a few cross-national studies that test hypotheses deriving from the anthropological literature.

² As of the time of writing, this feature was not available and therefore a check box does not appear in Figure 1.

³ HRAF is adding cases regularly, so these numbers will change.

⁴ The first author does not consider “Galton’s Problem” to be as serious as some other cross-cultural researchers (see discussion in Ember and Ember, 2009, 89–91).

REFERENCES

- Barry, Herbert, and Leonora M Paxson. (1971) *Infancy and Early Childhood: Cross-Cultural Codes 2*. *Ethnology* 10:466–508.
- Divale, William Tulio. (1975) *Temporal Focus and Random Error in Cross-Cultural Hypothesis Tests*. *Cross-Cultural Research* 10:19–36.
- Ember, C. R., and M. Ember. (1992) *Resource Unpredictability, Mistrust, and War – a Cross-Cultural-Study*. *Journal of Conflict Resolution* 36:242–62.
- Ember, Carol R. (2012) *Human Relations Area Files*. In *Leadership in Science and Technology*, edited by William Sims Bainbridge, pp. 619–27: Sage Reference.
- . (2007) *Using the Hraf Collection of Ethnography in Conjunction with the Standard Cross-Cultural Sample and the Ethnographic Atlas*. *Cross-Cultural Research* 41:396–427.
- Ember, Carol R, and Melvin Ember. (2010) *Explaining Male Initiation Ceremonies: New Cross-Cultural Tests and a Catalytic Model*. *Journal of Cross-Cultural Psychology* 41:605–16.
- Ember, Carol R, Hugh Page, Jr., Timothy O'Leary, and M. Marlen Martin. (1992) *Computerized Concordance of Cross-Cultural Samples*. New Haven, CT: Human Relations Area Files.
- Ember, Carol R., and Melvin Ember. (2009) *Cross-Cultural Research Methods*. 2nd ed. Lanham: Altamira Press.
- . (2009) *Cross-Cultural Research Methods*. 2nd ed. Lanham, MD: AltaMira Press.
- Ember, Carol R., Marc Howard Ross, Michael L. Burton, and Candice Bradley. (1991) *Problems of Measurement in Cross-Cultural Research Using Secondary Data*. *Behavior Science Research* 25:187–216.
- Human Relations Area Files, inc. *Cross-Cultural Research Overview*.
- Lagacé, Robert O. (1979) *The Hraf Probability Sample: Retrospect and Prospect*. *Cross-Cultural Research* 14:211–29.
- Murdock, George P, and Diana O Morrow. (1970) *Subsistence Economy and Supportive Practices: Cross-Cultural Codes 1*. *Ethnology*:302–30.
- Murdock, George P, and Douglas R White. (1969) *Standard Cross-Cultural Sample*. *Ethnology* 8:329–69.
- Murdock, George P. (1962–1971) *Ethnographic Atlas*. *Ethnology* 1–10.
- Murdock, George P.; Ford, Clellan S.; Hudson, Alfred E.; Kennedy, Raymond; Simmons, Leo W.; Whiting, John W. M. (2008) *Outline of Cultural Materials*. 6th edition with modifications ed. New Haven, CT: Human Relations Area Files.
- Murdock, George Peter. (1981) *Atlas of World Cultures*. University of Pittsburgh Pre.
- . (1957) *World Ethnographic Sample*. *American Anthropologist* 59:664–87.

- Naroll, Raoul. (1967) The Proposed Hraf Probability Sample. *Cross-Cultural Research* 2:70-80.
- Naroll, Raoul, and Richard G Sipes. (1973) A Standard Ethnographic Sample. *Current Anthropology* 14:111-40.

Table 1. Number of cross-cultural studies (with 10 or more cultures) by decade

Years	No. of Cross-Cultural Studies	Using HRAF in any form	Using SCCS
1889-1949	11 in six decades (average <2)	n.a.	n.a.
1950-1959	18	4	n.a.
1960-1969	136	35	n.a.
1970-1979	233	60	26
1980-1989	163	46	77
1990-1999	74	14	48
2000-2009	129	30	66

Table 2. SCCS cases not yet in eHRAF World Cultures as of June 2016; those with a HRAF ID are in the microfiche/paper Collection of Ethnography

SCCS ID	SCCS Name	SCCS Date	Community Name	District Name	HRAF ID	HRAF Name	Doc. Nos. in Paper/Fiche	eHRAF Name	eHRAF Doc. Matches
---------	-----------	-----------	----------------	---------------	---------	-----------	--------------------------	------------	--------------------

9	Hadza	+1930							
10	Luguru or Waluguru	+1925		Morogoro District					
15	Banen or Banyin	+1935							
25	Fulani	+1951							
27	Massa	+1910	Yagoua						
29	Fur	+1880	Jebel Marra						
30	Otoro	+1930	Nuba Hills						
32	Mao	+1939							
38	Bogo	+1855							
39	Barabra	+1900							
44	Hebrews	-0621		Judea					
45	Babylonians	-1750	Babylon						
50	Basques	+1934	Vera de Bidasoa	Spain	EX08	Basques	no direct matches	Basques	no direct matches
54	Russians	+1955	Viriatio Village		RF01	Great Russia	no direct matches		
56	Armenians	+1843	Erevan		RJ01	Armenia	no direct matches		
59	Punjabi	+1950	Mohla Village						
70	Lakher	+1930							
72	Lamet	+1940							
74	Rhade	+1962	Ko-Sier Village						
86	Badjau	+1963		Tawi-Tawi					
88	Tobeloese	1900							
93	Kimam	+1960	Bamol						
101	Pentecost	+1953	Bunlap						
102	Mbau Fijians	+1840		Bau Chiefdom					
103	Ajie	+1845		Neje Chiefdom					
107	Gilbertese								
113	Atayal	+1930							
114	Chinese	+1936	Kaixiangang Village		AF01	China	14		
120	Yukaghir	+1850							
130	Eyak	+1890							

133	Twana	+1860							
139	Kutenai	+1890							
143	Omaha	+1860			NQ12	Dhegiha	1,2,4,5	Omaha	1,2,4,5
144	Huron	+1634							
146	Natchez	+1718							
153	Aztec	+1520	Tenochtitlan		NU07	Aztec	2,3,4,5,6,8,9,10, 11,12,13,14,15, 16,17,18,19,20, 21,22,23,24,25, 26,27,28,29,30, 31,32,33,34,35	Aztec	2,3,4,5,6,8,13,14,15,16,17,18,19, 20,21,22,24,25,26,28,29,30,33, 34
154	Popoluca	+1940	Soteapan						
155	Quiche	+1930	Chichicastenango						
170	Amahuaca	+1960							
178	Botocudo	+1884							
182	Lengua	+1889							

Table 3. Two crosstabulations comparing Full SCCS to HRAF partial SCCS

GATHERING [Proportion of economic activity] by HUNTING [Proportion of economic activity]

Full SCCS - Observed Data

HUNTING [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
GATHERING [Proportion of economic activity]				
0 - 5%	102	25	10	137
6 - 15%				
16 - 25%	8	16	8	32
26 - 35%				
36 - 45%	1	11	5	17
46 - 55%				
56 - 65%				
66 - 75%				
76 - 85%				
Totals	111	52	23	186

HRAF SCCS - Observed Data

HUNTING [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
GATHERING [Proportion of economic activity]				
0 - 5%	73	14	6	93
6 - 15%				
16 - 25%	6	13	7	26
26 - 35%				
36 - 45%	1	10	3	14
46 - 55%				
56 - 65%				
66 - 75%				
76 - 85%				
Totals	80	37	16	133

Full SCCS - Contributions to X²

HUNTING [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
GATHERING [Proportion of economic activity]				
0 - 5%	5	4.6	2.8	137
6 - 15%				
16 - 25%	5.8	5.5	3.1	32
26 - 35%				
36 - 45%	7.3	8.2	2.7	17
46 - 55%				
56 - 65%				
66 - 75%				
76 - 85%				
Totals	111	52	23	186

HRAF SCCS - Contributions to X²

HUNTING [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
GATHERING [Proportion of economic activity]				
0 - 5%	5.2	5.4	1.9	93
6 - 15%				
16 - 25%	5.3	4.5	3.0	26
26 - 35%				
36 - 45%	5.6	9.5	0.3	14
46 - 55%				
56 - 65%				
66 - 75%				
76 - 85%				
Totals	80	37	16	133

X² = 45.4 with 4 degrees of freedom. p < .05

X² = 41.8 with 4 degrees of freedom. p < .05

BETROTHAL by AGRICULTURE [Proportion of economic activity]

Full SCCS - Observed Data

AGRICULTURE [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
BETROTHAL				
Bride- Price or - Wealth	22		49	71
Dowry	19		5	24
Totals	41		54	95

HRAF SCCS - Observed Data

AGRICULTURE [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
BETROTHAL				
Bride- Price or - Wealth	16		36	52
Dowry	14		3	18
Totals	30		40	70

Full SCCS - Contributions to X²

AGRICULTURE [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
BETROTHAL				
Bride- Price or - Wealth	2.4		1.8	71
Dowry	7.2		2.8	24
Totals	41		54	95

HRAF SCCS - Contributions to X²

AGRICULTURE [Proportion of economic activity]	0 - 5% 6 - 15%	16 - 25% 26 - 35%	36 - 45% 46 - 55% 56 - 65% 66 - 75% 76 - 85%	Totals
BETROTHAL				
Bride- Price or - Wealth	1.7		1.3	52
Dowry	5.1		3.2	18
Totals	30		40	70

X² = 16.3 with 1 degree of freedom. p < .05

X² = 11.4 with 1 degree of freedom. p < .05

Figure 1: Shows a portion of the table from HRAF's home page with matches to cultures in eHRAF World Cultures and permalinks to the matching documents

Show entries Search:

SCCS ID	SCCS DATE	SCCS NAME	EHRAF WORLD CULTURES NAME	OWC CODE	EHRAF DOCUMENTS
1	+1860	Nama or Namaqua	Khoi	FX13	Hoernlé, 1918 Hoernlé, 1923 Hoernlé, 1925 Schultze, 1907
2	+1950	Kung	San	FX10	Marshall, 1957 Marshall, 1957 Marshall, 1958 Marshall, 1959 Marshall, 1961 Marshall, 1962 Marshall, 1965 Marshall, 1976 Thomas, 1959
3	+1895	Thonga or Bathonga	Tsonga	FT06	Junod, 1927 Junod, 1927

Figure 2. How to select SCCS societies in eHRAF World Cultures

(next page)

Region and Culture Results

Finished. Found 27266 paragraphs in 2735 documents in 300 cultures

Search Query:

[Show](#) / [Hide](#)

✳ Narrow Results by Subsistence Type and Sample

By Subsistence Type

- Hunter-gatherers
- Primarily Hunter-gatherers
- Pastoralists
- Horticulturalists
- Intensive Agriculturalists
- Agro-pastoralists
- Other Subsistence Combinations
- Commercial Economy

By Sample

- Probability Sample Files (PSF)
- Standard Cross Cultural Sample (SCCS)
- Simple Random Sample (SRS)

▶ [Africa](#) (3221 paragraphs in 323 documents in 32 cultures)

Figure 3. SCCS Present/Absent Cases in HRAF Collection (133 present of 186).

(Note: figure could be improved by banding to show inclusiveness in regions 1–6)

1	2	3	4	5	<input type="checkbox"/>	7	8	<input type="checkbox"/>	<input type="checkbox"/>	11	12	13	14	<input type="checkbox"/>
16	17	<input type="checkbox"/>	19	20	21	22	23	<input type="checkbox"/>	<input type="checkbox"/>	26	<input type="checkbox"/>	28	<input type="checkbox"/>	<input type="checkbox"/>
31	<input type="checkbox"/>	33	34	<input type="checkbox"/>	36	37	<input type="checkbox"/>	<input type="checkbox"/>	40	41	42	43	<input type="checkbox"/>	<input type="checkbox"/>
46	47	48	49	<input type="checkbox"/>	<input type="checkbox"/>	52	53	<input type="checkbox"/>	55	<input type="checkbox"/>	57	58	<input type="checkbox"/>	60
61	62	63	64	65	66	67	68	69	<input type="checkbox"/>	71	<input type="checkbox"/>	73	<input type="checkbox"/>	75
76	77	<input type="checkbox"/>	79	80	81	82	83	84	85	<input type="checkbox"/>	87	<input type="checkbox"/>	89	90
91	92	<input type="checkbox"/>	94	95	96	97	98	<input type="checkbox"/>	100	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	104	105
106	<input type="checkbox"/>	108	109	110	111	112	<input type="checkbox"/>	<input type="checkbox"/>	115	116	117	118	119	<input type="checkbox"/>
121	<input type="checkbox"/>	123	124	125	126	127	<input type="checkbox"/>	129	<input type="checkbox"/>	<input type="checkbox"/>	132	<input type="checkbox"/>	134	135
136	137	138	<input type="checkbox"/>	140	<input type="checkbox"/>	142	143	<input type="checkbox"/>	145	<input type="checkbox"/>	147	148	149	150
151	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	156	157	158	159	160	161	162	163	164	165
166	167	168	169	<input type="checkbox"/>	171	172	173	174	175	176	177	<input type="checkbox"/>	<input type="checkbox"/>	180
181	<input type="checkbox"/>	183	184	185	186									

