Article

AnthroTools: An R Package for Cross-Cultural Ethnographic Data Analysis

Benjamin Grant Purzycki[1] and Alastair Jamieson-Lane[2]

Abstract

As large-scale collaborative, cross-cultural ethnographic research becomes easier and easier to realize, certain ethnographic methods and analyses should be correspondingly more available, inviting, and accommodating. We have therefore created AnthroTools, a package for the free, open-source language R, with a variety of tools and functions suitable for both multi-factor free-list analysis and Bayesian cultural consensus modeling. Free-list data elicitation is a simple technique for ethnographic research. However, especially for cross-cultural free-list data, background preparation is considerable and often requires specific software. In addition, although current cultural consensus analysis tools offer very sophisticated analyses, they also either require specialized software or have computationally taxing methods. AnthroTools expedites these techniques, rapidly performs diagnostics, and prepares data for further analysis. In this article, we briefly discuss what this package offers cross-cultural researchers and provide basic examples of some of its functions.

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]University of British Columbia, Vancouver, Canada

Corresponding Author:
Benjamin Grant Purzycki, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany.
Email: benjamin_purzycki@eva.mpg.de

Introduction

Analytical tools in cultural research often suffer from the need for a considerable amount of processing, cleaning, and preparation before analyses. Although extant free software such as ANTHROPAC/UCINET (Borgatti, 1996; ANTHROPAC henceforth) is ready-made for consensus and content domain analyses, such software often requires data to be organized in ways not suitable for integration with greater data sets and requires navigating software-specific interfaces that are not readily transferable to other programs. Because of this, we created AnthroTools (Jamieson-Lane & Purzycki, 2016), a simple package for use in R (R Core Team, 2012) to facilitate data management and analysis for social scientists interested in cross-cultural research. Currently, AnthroTools is equipped with two primary modes of analysis: (a) free-list data analysis and (b) cultural consensus analysis, both of which are foundational for cross-cultural and ethnographic research (see Bernard, 2011; Handwerker, 2001). In this article, we introduce the theory and practice of both methods of analysis, discuss AnthroTools in the context of other tools currently available, and provide walkthrough examples of the package's utility to perform said analyses. We refer readers to the package's help menu (run `help ["AnthroTools"]`) and website (https://anthrotools.wordpress.com) for further documentation, prefabricated code for analysis, walkthrough examples, and an archive for updates. There, users can also submit suggestions, report problems, and point to other packages for further analysis.

Free-List Tasks

Free-list data elicitation techniques are invaluable for assessing the content and structure of explicit human thought (Gravlee, 1988; Quinlan, 2005; Romney & D'Andrade, 1964; Schrauf & Sanchez, 2008; Smith, 1993; Smith & Borgatti, 1997; Smith, Furbee, Maynard, Quick, & Ross, 1995; Thompson & Juan, 2006). Free-list tasks generate data that—when appropriately analyzed—readily approximate our mental models of any given domain. Shared models and their content are "cultural models" (D'Andrade, 1981; Goodenough, 1957).

In these tasks, participants list objects in a given domain. At its core, the task lends itself to accounting for the ubiquity of specific items, but also to concept salience (i.e., its cognitive accessibility and/or importance; see below) across the minds of individuals. There are also many other useful accompanying methods to free-listing such as ranking, evaluation, and sorting (Brewer, 2002). Free-lists are useful as a preliminary step toward more focused research efforts such as scale designs and pile sorting (Bernard, 2011; Brewer, 2002;

Handwerker, 2001). Moreover, the presence, absence, or co-occurrence of listed items, item frequencies, and/or salience scores may also be used as dependent or independent variables in targeted studies (e.g., Purzycki, 2016; Schrauf & Sanchez, 2008) and thus may serve as an end in themselves.

Researchers have regularly used free-lists for nearly half a century. Some domains of inquiry using this technique have been the animals people know (Henley, 1969), color terms in English (Smith et al., 1995) and Estonian (Sutrop, 2001), wildflowers (Robbins & Nolan, 1997), plants that treat intestinal worms (M. B. Quinlan, Quinlan, & Nolan, 2002), things that people need to live a good life (Dressler, Borges, Balieiro, & dos Santos, 2005), the kinds of fruits they know (Bernard & Ryan, 2009; Hough & Ferraris, 2010), things that people think gods like and dislike (Purzycki, 2011, 2016; Purzycki et al., 2016), dishes of food across social classes in Argentina (Libertino, Ferraris, Osornio, & Hough, 2012), models of what constitutes good or bad people (Purzycki, 2016), and many, many others.

Free-list tasks are quite simple to execute, and provided that questions are relevant, they are virtually effortless for participants from all walks of life to answer. Researchers can limit the number of items listed, or ask participants to exhaust their available knowledge. In addition to responses, recording the order in which participants offer them is also important, as it lends itself to calculating cognitive salience or the accessibility of specific items. Items listed earlier are typically more salient and often more ubiquitous within a sample (Romney & D'Andrade, 1964). As discussed below, free-list data do require more backend processing than, for instance, scale-based quantitative data. However, the task has the benefits of not requiring forced responses (a common concern), it is a naturalistic method and easy for samples who have difficulty with scales, and it is quite easy to recode data using whatever scheme a researcher wishes to use thus facilitating cross-cultural comparison. The task is therefore ideal for navigating the relationship between emic and etic models of cultural information (Harris, 1976; Headland, Pike, & Harris, 1990; Pike, 1967).

Analysis

Despite the simplicity of the method, free-list data lend themselves to a wide variety of analyses. Aside from presence or absence of list items or overall item frequencies, we can also calculate item salience. As mentioned above, item order indicates the salience of a specific item for individuals; earlier listed items are taken to be more accessible components of mental models than those listed later. For example, if we were to ask someone to list things he or she thinks of when we say "dog," they might say "cat" first. If so, then

"cat" has a salience score of 1. Individual salience is simply the *inverse* order of items listed divided by the total number of items listed. So, if he or she listed "Labrador" third in a list of 20 items, its salience score is 0.9 (18 / 20 = 0.9). If the participant listed "purple" last, "purple" gets a salience score of 0.05 (1 / 20 = 0.05). Equation 1 is the standard procedure for calculating item salience for individuals:

$$\text{Item salience} = (n+1-k)/(n) \tag{1}$$

where *n* is the total number of items listed by an individual and *k* is the order in which an item was listed (e.g., *k* = 3 for "Labrador" in the above example).

Cultural models, however, consist of clusters of items with the highest frequency and often highest mean salience in a sample. In other words, widely shared components of mental models are the cultural content of human thought. Group salience is referred to as "Smith's S" and is simply the group mean of item salience (Borgatti, 1998; M. Quinlan, 2005; Smith, 1993; Smith et al., 1995). If, in a sample of 10, "cat" had an overall average salience score of 0.95, we can confidently infer that cat was listed often and early in lists. Cognitively speaking, we might then say that there was a very strong connection between the conceptual units for "dog" and "cat" (see Strauss & Quinn, 1997). If only one person in a sample of 10 listed "purple" and it had a salience score of 0.05, its Smith's S would be 0.005 (0.05 / 10 = 0.005) suggesting virtually no predictable conceptual links at all. Other such calculations abound (see Sutrop, 2001).

Although the method of analysis seems reasonably straightforward, doing such calculations "by hand" in a spreadsheet is notably time-consuming and requires reverse coding the order of items, calculating item salience by number of items people listed, repeating the process for each individual, and regularly sorting and resorting the data. In response to this, some dedicated people developed software to better facilitate data processing.

Current Software

The standard software for analyzing free-list data is ANTHROPAC (Borgatti, 1996; Borgatti, Everett, & Freeman, 2002) or FLAME (Pennec, Wencélius, Garine, Bohbot, & Raimond, 2016). ANTHROPAC requires that one organize free-list data in the following format in a text file (Table 1):

For multi-site projects with large sample sizes, creating this input method in a text format can be time-consuming and is not readily compatible with otherwise standard data storage formats such as spreadsheets, which better

Table 1. Generic Example of Free-List Data for Use in ANTHROPAC.

```
# informant1
item1
item2
item3
item4
item5
item6
# informant2
item1
item2
item3
item4
item5
item6
# informant3
item1
item2
item3
item4
item5
item6
```

facilitate data integration. ANTHROPAC in particular includes no capacity for multi-factor free-list data analysis. An alternative, FLAME (Pennec et al., 2016), works directly in Excel, has a slightly different data entry requirement, but does allow multi-factor analyses and has very useful plot and table functions. However, if one does not have or is not inclined to use for-pay software, uses a dated version of Microsoft Office, and/or has a very large data set, then he or she must resort to other alternatives (see Borgatti, 2015, for review). And finally, integrating outputs with greater data sets (e.g., demographics) requires the use of further programs.

One might also organize free-list data in a spreadsheet with the format detailed in Table 2. This method is helpful for conducting cross-cultural work in teams, as it is intuitive for those who typically work in a spreadsheet format and spares researchers the effort of creating long text files. This method is also helpful for creating quick tables with spreadsheet software (e.g., "pivot" or "pilot" tables in Excel or Calc, respectively). However, subsequent calculations (e.g., of item salience) or transformations (e.g., dichotomizing presence and absence) can be time-consuming and prone to error, particularly if one has

Table 2. Generic Example of Free-List Text File for Use in Spreadsheet.

Participant ID#	Order listed	Item listed
1	1	item1
1	2	item2
1	3	item3
2	1	item1
2	2	item2
2	3	item3
3	1	item1
3	2	item2
4	1	item1
4	2	item2
4	3	item3
4	4	item4

a large data set with many participants and groupings. If researchers are not inclined or able to write and maintain macros that often require updating with subsequent versions of software, these tasks can be especially tedious and frustrating if specific functions do not work or require debugging. In addition, macros for one office suite rarely transfer to others making the analyses difficult to achieve without the right software, and hindering transfer of relevant macros between researchers.

AnthroTools overcomes these limitations and allows researchers to (a) quickly analyze free-list data; (b) convert them into various data sets that are more immediately usable for merging with other data and subsequent analyses; (c) take advantage of the versatility of the free, open-source language R (R Core Team, 2012) that has a plethora of extant packages that handle typical analytical accompaniments to free-list tasks (e.g., multi-dimensional scaling, correlational matrices, principal components analysis, regression, etc.). R has full graphics capabilities and handles a variety of data file formats. AnthroTools also (d) includes built-in help files with illustrations and examples for ease in use. And, in the event that researchers already have data in the ANTHROPAC format (.txt files), it can readily convert data in this format into a standardized spreadsheet for immediate use.

As described in more detail below, we also created a variety of options for transforming free-list data into useful tables for further integration and analysis (e.g., using free-list data as a dependent or independent variable, running principal components analysis, etc.). These options include (a) tables with binary data indicating whether or not an individual listed a specific item

(ideal for logistic regression), (b) tables of the frequency of items (ideal for free-list data that are subsequently coded using a standardized coding rubric), (c) tables that report the sum of salience for each individual item per participant, (d) a table including the order values of listed items, and (e) a table reporting the maximum salience for each item per individual. Subsequent generation of participant × participant or item × item matrices (e.g., for use with multi-dimensional scaling) are already built into R's framework. Virtually every other step from this stage in analysis is already available in R's many other packages and default functions.

Cross-Cultural Free-Listing

Researchers regularly use free-lists to rapidly make sense of cultural models within a sample. However, cross-cultural researchers may wish to collect data in the same or similar cultural domains from multiple sites or from multiple groups (e.g., Purzycki et al., 2016). Indeed, we developed the software to handle such cross-site data quickly and efficiently. For example, if your theory predicts that Item X will be more salient in Group Y than in Group Z, then grouping salience scores is necessary. If you predict that a certain class of items is going to be more prevalent in one group or another, again, grouping is imperative. If some domains should contain very similar items regardless of culture, once again, chunking salience and frequencies by group is required to effectively examine this. AnthroTools can chunk free-list data by any factor you wish (e.g., gender, religious group, cultural site, age group, etc.). Simply identifying how to chunk calculations will immediately generate cross-cultural sets, reports, and matrices that are ready for global and local analyses helpful for team-based cross-cultural research management. To illustrate, we offer a brief walkthrough of a free-list analysis below. Like many R packages, AnthroTools includes built-in examples for all of its primary functions. We highlight some of the main features in this article, but more functions and details are available within the package and in the package documentation on the website.

Walkthrough Example

To install AnthroTools, run the following code in R:

```
install.packages("devtools")
library("devtools")
install_github('alastair-JL/AnthroTools')
library(AnthroTools)
```

Table 3. Truncated Data Set of Freely Listed Fruits.

Subj	Order	Code
1	1	pear
1	2	orange
1	3	apple
2	1	apple
2	2	apple
2	3	strawberry
3	1	banana

The only additional package upon which AnthroTools relies is "devtools" (Wickham & Chang, 2016).

Salience analysis. To illustrate, let us assume that you have data from people who freely listed the kinds of fruits they knew (Bernard & Ryan, 2009). After installing and loading the package, simply running

```
data(FruitList)
```

will call up the sample free-list data we created. To view the data set (Table 3), run:

```
View(FruitList)
```

Note that in this data set, there are only three variables: "Subj" is the participant ID number, "Order" is the order in which participants listed items, and "CODE" is the data point. In total, there are 20 participants.

The CalculateSalience command generates each item's salience score and creates a new column for this score. We will create a new object called "FL" that will be this new data set:

```
FL <- CalculateSalience(FruitList)
View(FL)
```

Table 4 is a truncated version of this new set with "Salience" scores added as a new column:

The full syntax for CalculateSalience is as follows:

```
FL <- CalculateSalience(dataset, Order = "Order", Subj =
"Subj", CODE = "CODE", GROUPING = NA, Rescale = FALSE,
Salience = "SScore")
```

Table 4. Truncated Data Set of Freely Listed Fruits After Calculating Salience.

Subj	Order	CODE	Salience
1	1	pear	1.0000000
1	2	orange	0.6666667
1	3	apple	0.3333333
2	1	apple	1.0000000
2	2	apple	0.6666667
2	3	strawberry	0.3333333

This allows you to call your variables whatever you want and associate them with the function's operations (e.g., if your participant ID variable is called "ID" in your data set, the code would be Subj = "ID" for that component).

The function "SalienceByCode" takes this new data set with the by-item salience scores and creates another output specifically designed to handle overall salience by item code (Table 5). Run the following:

```
FL.S <- SalienceByCode(FL, dealWithDoubles = "MAX")
View(FL.S)
```

This argument calculates by-item mean salience and the sum of salience scores. It calculates Smith's S by dividing the sum of salience scores by number of participants in the sample (again, in this case, $n = 20$; Borgatti, 1998; M. Quinlan, 2005; Smith, 1993; Smith et al., 1995).

As mentioned above, data points often get repeated due to individual responses or post hoc recoding. To avoid the subsequent inflation of Smith's S values, we created the "dealWithDoubles" argument to allow flexibility in handling such instances. On the default setting, the function will assume that no such cases arise. If there are such cases, R will report an error and encourage you to use the "dealWithDoubles" command. So, if you run the same script above, but delete the "dealWithDoubles" argument, you will see the error notifying you of repeated items (notice that Subject 2 lists "apple" twice).

Aside from DEFAULT, you also have the options MAX, MEAN, SUM, and IGNORE. MAX indicates that you want the computer to attend *only* to the first time a respondent lists a particular CODE, and ignore subsequent mentions (e.g., if someone lists "apple" twice, it only keeps track of its earliest listing). For MEAN, the computer determines each respondent's mean salience for repeated items and uses this value to calculate Smith's S.

Table 5. Salience Analysis of Free-List Items.

	CODE	MeanSalience	SumSalience	SmithsS
1	pear	0.695833	5.566667	0.278333
2	orange	0.644444	3.866667	0.193333
3	apple	0.758889	11.383333	0.569167
4	strawberry	0.592593	5.333333	0.266667
5	banana	0.731250	5.850000	0.292500
6	plum	0.811905	5.683333	0.284167
7	lemon	0.508333	2.033333	0.101667
8	peach	0.875000	1.750000	0.087500

For SUM, you are asking the computer to determine each respondent's total salience with respect to a given code. If this value is greater than "1," R will report an error identifying the source of the problem and recommend normalization (see below). IGNORE is merely a way of suppressing errors and is thus not recommended. The full syntax for SalienceByCode is as follows:

```
FL.S <- SalienceByCode(FL, Subj = "Subj", CODE =
"CODE", GROUPING = "GROUP", Salience = "Salience",
dealWithDoubles = "DEFAULT")
```

Tables for further analyses. As noted above, we created a variety of options for transforming free-list data into useful tables for further analyses (e.g., using free-list data as a dependent or independent variable, running factor analysis, etc.). The general syntax is as follows:

```
FLT <- FreeListTable(dataset, CODE = "CODE", GROUPING
= "GROUP", Salience = "Salience", Subj = "Subj",
tableType = "DEFAULT")
View(FLT)
colSums(FLT)
```

Currently, there are five types of tables: "PRESENCE," "SUM_SALIENCE," "MAX_SALIENCE," "HIGHEST_RANK," and "FREQUENCY." "PRESENCE" converts data into a "1" if participants mentioned the specified code and "0" if they did not (Table 6). This might be helpful for logistic regressions where your dependent variable is whether

Table 6. Truncated Table of Whether or Not Participants Listed a Specific Item (the "PRESENCE" Argument).

Subject	Pear	Orange	Apple	Strawberry	Banana	Plum	Lemon	Peach
1	1	1	1	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	1	1	0	0	0
4	0	1	1	0	1	1	0	0
5	0	0	1	1	0	0	0	0
6	0	0	1	0	0	0	1	0
7	0	0	0	0	1	1	1	0
8	1	0	1	1	0	0	0	0
9	0	0	1	0	0	1	0	0

or not someone listed a specific type of item or to generate proximity matrices. If you use "SUM_SALIENCE," then you will get the total salience each person has associated with each code. If you use "MAX_SALIENCE," then you will get the maximum salience, that is, the salience of the code the first time it was mentioned. "HIGHEST_RANK" creates a Participant × Item matrix of the order in which each item initially appeared for each individual. These are especially useful for multi-dimensional scaling to craft visual mental maps of items. If you specify "FREQUENCY," then you will get a count of how often each code was mentioned by each person. This might be useful in general linear models, especially when your data consist of relatively frequently repeated items. Quick summations for each column can be made with R's built-in colSums() command. To illustrate, if we run the following code on the FL data set, we get a table with binary presence/ absence data for lists.

```
FLT <- FreeListTable(FL, CODE = "CODE", Salience =
"Salience", Subj = "Subj", tableType = "PRESENCE")
View(FLT)
```

Again, such tables may be useful in and of themselves or in tandem with other data sets of the same participants to make predictions about or using frequency and salience. Moreover, cross-cultural variation in item salience is also of importance and interest. If, for instance, you conduct the same free-list task among multiple, culturally disparate groups, chunking salience by group may be necessary. We now illustrate this procedure in AnthroTools.

Table 7. Truncated Salience Analysis of Free-List Items Using the GROUPING Argument.

GROUPING	CODE	MeanSalience	SumSalience	SmithsS
MAINLAND	plum	0.8333	5.0000	0.2941
ISLAND	plum	0.7729	6.1833	0.4756
MOON	plum	0.5833	1.7500	0.1250
MAINLAND	strawberry	0.6875	2.7500	0.1618
ISLAND	strawberry	0.5000	2.0000	0.1538
MOON	strawberry	0.4722	1.4167	0.1012
MAINLAND	pear	0.5833	2.9167	0.1716
ISLAND	pear	0.5733	2.8667	0.2205
MOON	pear	0.3500	1.4000	0.1000

Grouping salience calculations. Grouping allows salience analysis to be carried out on individuals belonging to different groups (e.g., different field sites). This option is optimal for cross-cultural or any comparative study where participants are grouped by any given factor. Again, simply identifying which factor you wish to consider by using the GROUPING argument will perform the full task. Subject code overlap between groups is permitted, and salience calculations can be done on a group-by-group basis for each item code. The following includes syntax for the toy example included in the package. You will see that the final output is a table calculating salience by item, by group.

```
data(WorldList)
WL <- CalculateSalience(WorldList, GROUPING =
"GROUPING")
WL.S <- SalienceByCode(WL, dealWithDoubles = "MAX",
GROUPING = "GROUPING")
View(WL.S)
```

The output will look something like Table 7. The GROUPING variable here divides the analysis by sample or cultural group—the mainlanders, people from the island, and some people from the moon. Note that we sorted by CODE. Mean salience, sum of salience scores, and Smith's S are all calculated based on within-group sample sizes (we truncated the table and the decimal places). The FreeListTable function also works with the grouping argument, so one may generate similar data sets (binary, frequency, etc.) with the grouping variables intact.

Cleaning free-list data. Free-list data and their analyses face common problems, and we designed AnthroTools to be able to address them (`see help [CleanFreeList]`). The first function is for error checking and removal. In the event that you wish to streamline various error correction associated with such data, you can use this function to do so. This cleaning function allows you to (a) remove participants who have bad data (e.g., missing order values, repeated items, etc.), (b) correct order items (e.g., if a participant's order numbers are 1, 2, 8, 9, the function will convert it to 1, 2, 3, 4), and (c) expunge such cases if order is missing. The package will also report inconsistencies and errors in data (e.g., if multiple participants have the same ID number, if the order of items for individuals does not start at 1, if you have multiples of the same order number for individuals, etc.).

The second addresses variation in how many items participants list. In open-ended free-lists, some individuals will list many items thus swamping salience scores. AnthroTools can rescale and normalize salience scores across individuals. The third cleaning function addresses salience inflation. And again, as is often the case, data points get repeated; participants sometimes list specific items more than once. Alternatively, if free-list data are subsequently recoded using a prefabricated scheme, data will often appear to repeat (e.g., someone says "cat" and "tiger" and you wish to lump these together in your "feline" category). Repeated items can inflate Smith's S values in automated analyses. AnthroTools handles such items with multiple types of corrections including calculations based on earliest listed repeated items and the mean salience of such cases.

Cultural Consensus Analysis

Cultural consensus analyses (Anders & Batchelder, 2012, 2013; Anders, Oravecz, & Batchelder, 2014; Caulkins, 2004; Miller, Kaneko, Bartram, Marks, & Brewer, 2004; Oravecz, Anders, & Batchelder, 2013; Oravecz, Vandekerckhove, & Batchelder, 2014; Romney, Weller, & Batchelder, 1986) were designed to assess inter-informant agreement, expertise, and cultural knowledge. It is especially suitable for multiple-choice and dichotomous yes/no questions, where the interviewer does not necessarily know the correct answer or where the notion of a "correct" answer may itself be misguided. The traditional form of analysis searches for correlations between respondents' answers and uses these to infer the level of "expertise" of each respondent, before going on to determine which answer is considered correct by the majority of respondents (weighted such that more attention is paid to respondents

with greater expertise). By conducting consensus analysis with multiple groups, a researcher is able to systematically determine similarities and differences in what answers are deemed "correct."

Such analysis might be applicable to such questions as "What is the name of this plant?" or "Out of these five examples, which is most likely poisonous?" or "Is your deity omniscient?" Indeed, consensus analysis has been useful in a variety of domains ranging from evaluations of which material items indicate being a successful individual (Dressler, 1996) and social views of the wives of miners from Yorkshire (Caulkins & Hyatt, 1999) to assessments of traffic medians and safety (Kim, Donnell, & Lee, 2008) and European views on obesity (Ulijaszek, 2007).

The analysis depends on several main assumptions (Hruschka & Maupin, 2013; Romney et al., 1986), but the two most crucial are that questions have a single "correct" answer and that all questions are in the same domain. The traditional form of consensus analysis rests on the assumption that you are asking multiple-choice questions, and that for any given question, participants know the answer to the question with some probability p and guess the answer with some probability $1 - p$. Here, p is expected to vary between participants depending on their "competence" but does not vary between questions; it is assumed that all questions are of similar difficulty and within the same domain of knowledge (see more below).

Successful consensus analysis will produce two major outputs. First is an answer key for your survey items; it will tell you what the "correct" answer is for each question based on a variety of calculations. Second, it generates competence scores for each individual, indices that approximate the degree to which each participant resembles the cultural model.

The Logic of Cultural Consensus Analysis

This section explains the traditional logic behind cultural consensus theory and analysis, followed by sections reviewing extant software and a walk-through example of consensus analysis in AnthroTools.

The first step in the analysis is to estimate the value of p—the probability of knowing the right answer—associated with each individual. This is done by finding the correlation between individuals' answers, **M**, and then accounting for the probability that two individuals might pick the same answer by chance, resulting in a new matrix **M***, which is designed to represent the rate at which each pair of individuals select the same answer because they both knew the correct answer. Entries of **M*** are found using the following equation:

$$M^*_{i,j} = \left(LM_{i,j} - 1\right)/\left(L-1\right) \qquad (2)$$

Here, L represents the number of possible answers to a given question.

Assuming all questions and individuals are relatively independent of one another, this value (the rate of both knowing the answer) should be equal to the product of the p values for the two individuals.[1] Thus, for every pair of individuals, we have the following equation:

$$M^*_{i,j} = p_i p_j \qquad (3)$$

Here, we refer to the i,jth entry of the matrix, and the values of p refer to the competence of subjects i and j.

In general, this system will result in more equations than unknown variables, and thus cannot be solved exactly. However, assuming that all the assumptions of consensus analysis have been met, the set of equations should be *approximately* soluble. The AnthroTools package does this using what we will call the "Comrey Iteration" (Comrey, 1962, see more below).[2]

Once we have determined the competence scores of each individual, the "correct" answer to each question can be inferred from the survey by using Bayes' theorem:

$$P\left(A|B\right) = \frac{P\left(B\,|\,A\right)P\left(A\right)}{P\left(B\right)} \qquad (4)$$

In the particular case of consensus analysis, this translates to "the probability of answer A being correct for Question 1 given our current survey results is equal to the probability of seeing *this particular* series of answers *given* that A was correct, multiplied by the prior probability that A was correct, and divided by the probability of seeing this particular series of answers (without knowing anything about the answer)."[3]

Walkthrough Example

Basic consensus analysis. AnthroTools automates consensus analysis with the "ConsensusPipeline" function. The following is a walkthrough of our toy example. First, call up the data set:

```
data(ConsensusTestData)
View(ConsensusTestData)
```

Table 8 represents the example data set:

Table 8. Data Set Example of Person × Question Matrix for Cultural Consensus Analysis (Possible Answer Options Vary Between 1 and 3).

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Person 1	2	1	2	2	1	2	3	1
Person 2	3	2	1	3	1	1	2	1
Person 3	2	2	1	1	3	2	3	2
Person 4	1	2	3	1	2	1	3	1
Person 5	1	3	1	3	1	2	2	2
Person 6	2	2	1	2	3	1	2	2
Person 7	3	2	1	1	1	1	2	1
Person 8	3	2	3	1	2	1	2	3

To run the consensus analysis, simply invoke the function "ConsensusPipeline" and indicate the number of total possible answers to your multiple-choice key (in this case, it is 3):

```
Results <- ConsensusPipeline(ConsensusTestData, 3)
Results
```

These results contain the calculated answers, the competence assumed for each person, the actual test scores found for each person (assuming the answers calculated are correct), and the probability distribution for each of the answers. Note that in general, we are primarily interested in the answers calculated, but looking at the other results can be useful for informing you if your input has led to strange results (e.g., negative probabilities). Output 1 has the following elements: $Answers, $Competence, $origCompetence, $TestScore, $Probs, and $reportback.

```
$Answers

Question 1 Question 2 Question 3 Question 4
Question 5 Question 6 Question 7 Question 8


        3          2          1          1
        1          1          2          1

$Competence

Person 1 Person 2 Person 3 Person 4
Person 5 Person 6 Person 7 Person 8
```

```
0.000000000 0.730098933 0.007199317 0.269227910
0.026867347 0.245326389 0.973873955 0.513715567
```

$origCompetence

```
[1] -0.284995564 0.730098933 0.007199317
0.269227910  0.026867347  0.245326389  0.973873955
0.513715567
```

$TestScore

```
Person 1 Person 2 Person 3 Person 4 Person 5
Person 6 Person 7 Person 8

    0.250    0.875    0.375    0.500    0.375
    0.500    1.000    0.625
```

$Probs

```
Question 1 Question 2 Question 3 Question 4
Question 5 Question 6 Question 7 Question 8

1 0.0010616681 0.001750085 0.963055951 0.955268538
0.962671730 0.9979612622 0.0034738974 0.982874743

2 0.0004699594 0.996354875 0.006855928 0.007459002
0.030349790 0.0002476973 0.9955920386 0.001984112

3 0.9984683724 0.001895040 0.030088121 0.037272460
0.006978479 0.0017910405 0.0009340641 0.015141145
```

$reportback

[1] "We encountered 1 individuals with ''negative'' competence. We found 0 individuals with competence over one. The magnitude of the main factor was 1.4000155273283. The second factor's magnitude was 1.06401058316597, giving a ratio of 1.31579097941164."

$reportNumbers

```
[1] 1.000000 0.000000 1.400016 1.064011 1.315791
```

Output 1. Cultural consensus analysis output

The $Answers section provides the culturally "correct" answers provided by the function, whereas $Competence reports the estimated cultural competence for each individual. In this case, then, Persons 2 (≈ 0.73) and 7 (≈ 0.97) had the highest two competence and test scores, suggesting cultural expertise. $origCompetence is the competence score before the transformation into the [0, 1] range. Assuming, of course, that the answer key is correct, $TestScore reports test scores for each individual whereas $Probs is the probability that each answer is correct for a given question. For example, 0.9984683724 is the calculated probability that "3" is the correct answer to the first question, while 0.955268538 is the program's belief that "1" is the correct answer for question four.

We also provide various qualitative reports in the output. $reportback gives important feedback from the analyses. In the case of strange results, it may be worthwhile to examine whether or not the assumptions have been satisfied. The return value "reportback" has a written summary detailing such anomalies, including the competence scores that are out of bounds, along with the "Comrey Ratio"[4] used to determine whether it is likely that a single domain is present. "reportNumbers" yields the raw numbers from this summary, should you wish to analyze them.

In particular, for the above analysis, we notice that the "Comrey Ratio" is below 3. According to a standard "rule of thumb" ratio commonly referenced in the literature (e.g., Borgatti & Halgin, 2011; Weller, 2007), this ratio can be used to determine whether or not the assumption that all participants belong to a single unified culture (with a single correct answer key)—if this ratio is three or above, the assumption of a single unified culture is treated as valid; below this ratio, the assumption is considered suspect. However, in $reportback, the reported ratio was ≈ 1.32. This indicates that, according to this rule of thumb, the above survey cannot be assumed to come from individuals of a single culture, despite the fact that the data are computer-generated and were generated in a manner satisfying all assumptions of the model (single culture, independent answers, all questions in a single domain with equal difficulty).

Stress testing. To explore this phenomenon, we created the functions "GenerateConsensusData" and "ConsensusStressTest," the former of which can be used to simulate a single survey, whereas the latter can simulate a large number of surveys and then perform consensus analysis on each of them. As input, they both take a number of individuals (per survey), a number of questions, and a number of allowable answers (per question). The Stress Test also requires the number of surveys to simulate. In both cases, the user can also

choose to "lock" the competence of individuals setting all competence values of the group to the same value.

```
StressTestResults <- ConsensusStressTest(8, 8, 3,
5000,lockCompetence = 0.6)
```

These functions prove useful when attempting to determine whether a given result is "normal" in some sense. Again, in our test case example above, we found the Comrey ratio to be only 1.32—well below the 3:1 ratio recommended rule of thumb.

The stress test reports back a wide variety of statistics (please see the command `help(ConsensusStressTest)` for full details), but in particular, it can be determined what fraction of surveys resulted in a ratio greater than 3:1:

```
mean(StressTestResults[[4]]>3)
```

Upon running the above command ourselves, we found that for this size of survey, a mere 1.2% of all simulated surveys were able to pass the 3:1 rule of thumb. This suggests that the threshold currently recommended in the literature may need to be reconsidered—possibly taking into account survey size and number of participants. Alternatively, other statistical tests no longer dependent on this ratio of factor weightings may be more appropriate. For now, however, in the absence of stronger systematic methods for assumption checking, we offer the Consensus Stress Test as a tool for generating "typical" data for the researcher to compare against statistically.

Current Software

Like AnthroTools, ANTHROPAC conducts consensus analysis, but not in its updated Bayesian formulation. Oravecz et al. (2014) created a standalone program that offers this alternative for binary data, and Anders (2014) created the R package CCTpack for those who prefer to remain within the framework of R and handle a variety of data using contemporary Bayesian estimations. Both have a variety of very sophisticated applications. Our observations so far are that these more thorough methods spend significant amounts of time in their parameter estimation, particularly for large data sets, whereas the more approximate method provided in our package provides a significantly quicker result. In cases where simply determining the "correct" answers is required, and your data set is large enough to limit the issues caused by approximation, AnthroTools may thus prove useful as a quick and

streamlined exploratory measure (especially useful for field researchers), with the heavier tools offered by these sources used for validity testing and more advanced analytical techniques.

Conclusion

Our primary goals for developing this tool were to make free-list and consensus analyses easier to manage, readily accessible, more efficient, and immediately useable for field researchers interested in content domains. It should be especially helpful for researchers doing cross-cultural work, or working with large samples. It is our hope to inspire cultural anthropologists to use these powerful and useful methods who would otherwise avoid free-list data and cultural consensus analyses because of their backend work or mathematical esoterica. For those already using these tools, we hope this makes the process easier.

Notes

1. If I know the correct answer two third of the time, and you know the correct answer two third of the time, then we both know the correct answer four ninth of the time.
2. There is nothing inherent in Comrey's method of factor analysis to prevent competence scores above one or below zero. Furthermore, in going from M to $M*$, it is possible to introduce negative values, for example, if two individuals agree with one another *less* than would be expected by chance, then the method may infer that the two are in active disagreement, leading to unusual results.

3. In practice, once competence is known, it is easy to calculate P(survey results | answer is A), and all else being equal, it can be assumed that P(answer is A) is equal to P(answer is B), and so on, leading to P(answer is A) = 1/N (where N is the number of possible answers). The only remaining difficulty is determining P(survey results), but this we can avoid calculating simply by remembering that the probabilities of all possible answers must add up to one.

4. Here, we refer to the "Comrey Ratio" as the ratio between the primary and secondary vectors found when determining the competence scores from the agreement matrix. Previous explanations of consensus analysis frequently refer to this as an "eigenvalue" ratio (Borgatti & Halgin, 2011; Weller, 2007). To the best of our understanding, the eigenvalues of the agreement matrix are not the appropriate numbers to use in this context. Based on the context, the lengths of the factors found with the Comrey iteration appears to be the closest appropriate analogue (cf. Borgatti & Halgin, 2011). It is unclear whether previous authors were using the correct concept with the wrong label, or the correct label with the wrong concept. For further discussion of this difficulty, run help(ConsensusCaveats), although in either case, the "criterion" lacks strong justification.

References

Anders, R. (2014). *CCTpack: Cultural consensus theory applications to data* (R Package Version 1.4).

Anders, R., & Batchelder, W. H. (2012). Cultural consensus theory for multiple consensus truths. *Journal of Mathematical Psychology*, *56*, 452-469.

Anders, R., & Batchelder, W. H. (2013). Cultural consensus theory for the ordinal data case. *Psychometrika*, *80*, 151-181.

Anders, R., Oravecz, Z., & Batchelder, W. H. (2014). Cultural consensus theory for continuous responses: A latent appraisal model for information pooling. *Journal of Mathematical Psychology*, *61*, 1-13.

Bernard, H. R. (2011). *Research methods in anthropology* (5th ed.). Lanham, MD: AltaMira Press.

Bernard, H. R., & Ryan, G. W. (2009). *Analyzing qualitative data: Systematic approaches*. Thousand Oaks, CA: Sage.

Borgatti, S. P. (1996). *ANTHROPAC 4*. Columbia, SC: Analytic Technologies.

Borgatti, S. P. (1998). Elicitation techniques for cultural domain analysis. In J. J. Schensul, M. D. LeCompte, B. K. Nastasi, & S. P. Borgatti (Eds.), *The ethnographer's toolkit* (Vol. 3, pp. 115-151). Walnut Creek, CA: AltaMira Press.

Borgatti, S. P. (2015). Software review: FLAME (Version 1.1). *Field Methods*, *27*, 199-205.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Borgatti, S. P., & Halgin, D. S. (2011). Consensus analysis. In D. B. Kronenfeld, G. Bennardo, & V. C. de Munck (Eds.), *A companion to cognitive anthropology* (pp. 171-190). Malden, MA: Wiley-Blackwell.

Brewer, D. D. (2002). Supplementary interviewing techniques to maximize output in free listing tasks. *Field Methods*, *14*, 108-118.

Caulkins, D. D. (2004). Identifying culture as a threshold of shared knowledge: A consensus analysis method. *International Journal of Cross Cultural Management*, *4*, 317-333.

Caulkins, D. D., & Hyatt, S. B. (1999). Using consensus analysis to measure cultural diversity in organizations and social movements. *Field Methods*, *11*, 5-26.

Comrey, A. L. (1962). The minimum residual method of factor analysis. *Psychological Reports*, *11*, 15-18.

D'Andrade, R. G. (1981). The cultural part of cognition. *Cognitive Science*, *5*, 179-195.

Dressler, W. W. (1996). Culture and blood pressure: Using consensus analysis to create a measurement. *Field Methods*, *8*, 6-8.

Dressler, W. W., Borges, C. D., Balieiro, M. C., & dos Santos, J. E. (2005). Measuring cultural consonance: Examples with special reference to measurement theory in anthropology. *Field Methods*, *17*, 331-355.

Goodenough, W. H. (1957). Cultural anthropology and linguistics. In D. H. Hymes (Ed.), *Language in culture and society: A reader in linguistics and anthropology* (pp. 36-39). New York, NY: Harper & Row.

Gravlee, L. (1988). *The uses and limitations of free listing in ethnographic research*. Research Methods in Cognitive Anthropology. Retrieved from http://gravlee.org/ang6930/freelists.htm

Handwerker, W. P. (2001). *Quick ethnography*. Lanham, MD: AltaMira Press.

Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology*, *5*, 329-350.

Headland, T. N., Pike, K., & Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Newbury Park, CA: Sage.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, *8*, 176-184.

Hough, G., & Ferraris, D. (2010). Free listing: A method to gain initial insight of a food category. *Food Quality and Preference*, *21*, 295-301.

Hruschka, D. J., & Maupin, J. N. (2013). Competence, agreement, and luck testing whether some people agree more with a cultural truth than do others. *Field Methods*, *25*, 107-123.

Jamieson-Lane, A., & Purzycki, B. G. (2016). *AnthroTools: Some custom tools for anthropology* (R Package Version 0.8).

Kim, T.-G., Donnell, E. T., & Lee, D. (2008). Use of cultural consensus analysis to evaluate expert feedback of median safety. *Accident Analysis & Prevention*, *40*, 1458-1467.

Libertino, L., Ferraris, D., Osornio, M. M., & Hough, G. (2012). Analysis of data from a free-listing study of menus by different income-level populations. *Food Quality and Preference*, *24*, 269-275.

Miller, M. L., Kaneko, J., Bartram, P., Marks, J., & Brewer, D. D. (2004). Cultural consensus analysis and environmental anthropology: Yellowfin Tuna fishery management in Hawaii. *Cross-Cultural Research*, *38*, 289-314.

Oravecz, Z., Anders, R., & Batchelder, W. H. (2013). Hierarchical Bayesian modeling for test theory without an answer key. *Psychometrika*, *80*, 341-364.

Oravecz, Z., Vandekerckhove, J., & Batchelder, W. H. (2014). Bayesian cultural consensus theory. *Field Methods*, *26*, 207-222.

Pennec, F., Wencélius, J., Garine, E., Bohbot, H., & Raimond, C. (2016). *Flame v1.2- Free-List analysis under Microsoft Excel* (Software and English User Guide). Retrieved from https://www.researchgate.net/publication/261704624_Flame_ v12_-_Free-List_Analysis_Under_Microsoft_Excel_Software_and_English_ User_Guide

Pike, K. L. (1967). *Language in relation to a unified theory of the structure of human behavior*. The Hague, The Netherlands: Mouton.

Purzycki, B. G. (2011). Tyvan cher eezi and the socioecological constraints of supernatural agents' minds. *Religion, Brain & Behavior*, *1*(1), 31-45.

Purzycki, B. G. (2016). The evolution of gods' minds in the Tyva Republic. *Current Anthropology*, *57*, S88-S104.

Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., Xygalatas, D., Norenzayan, A., Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, *530*, 327-330.

Quinlan, M. (2005). Considerations for collecting freelists in the field: Examples from ethnobotany. *Field Methods*, *17*, 219-234.

Quinlan, M. B., Quinlan, R. J., & Nolan, J. M. (2002). Ethnophysiology and herbal treatments of intestinal worms in Dominica, West Indies. *Journal of Ethnopharmacology*, *80*, 75-83.

R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robbins, M. C., & Nolan, J. M. (1997). A measure of dichotomous category bias in free listing tasks. *Field Methods*, *9*, 8-12.

Romney, A. K., & D'Andrade, R. G. (1964). Cognitive aspects of English kin terms. *American Anthropologist*, *66*, 146-170.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*, 313-338.

Schrauf, R. W., & Sanchez, J. (2008). Using freelisting to identify, assess, and characterize age differences in shared cultural domains. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *63*, S385-S393.

Smith, J. J. (1993). Using ANTHOPAC 3.5 and a spreadsheet to compute a free-list salience index. *Field Methods*, *5*, 1-3.

Smith, J. J., & Borgatti, S. P. (1997). Salience counts—And so does accuracy: Correcting and updating a measure for free-list-item salience. *Journal of Linguistic Anthropology*, *7*, 208-209.

Smith, J. J., Furbee, L., Maynard, K., Quick, S., & Ross, L. (1995). Salience counts: A domain analysis of English color terms. *Journal of Linguistic Anthropology*, *5*, 203-216.

Strauss, C., & Quinn, N. (1997). *A cognitive theory of cultural meaning*. Cambridge, UK: Cambridge University Press.

Sutrop, U. (2001). List task and a cognitive salience index. *Field Methods*, *13*, 263-276.

Thompson, E. C., & Juan, Z. (2006). Comparative cultural salience: Measures using free-list data. *Field Methods*, *18*, 398-412.

Ulijaszek, S. J. (2007). Frameworks of population obesity and the use of cultural consensus modeling in the study of environments contributing to obesity. *Economics & Human Biology*, *5*, 443-457.

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods*, *19*, 339-368.

Wickham, H., & Chang, W. (2016). *devtools: Tools to make developing R Packages Easier* (R Package Version 1.12.0). Retrieved from https://CRAN.R-project.org/package=devtools

Author Biographies

Benjamin Grant Purzycki is currently a senior researcher at the Max Planck Institute for Evolutionary Anthropology (PhD, 2012, University of Connecticut). He engages in the cognitive, evolutionary, and ethnographic sciences of sociality and cultural variation with an emphasis on religion's role in mitigating local socioecological problems. He has conducted fieldwork in the Tyva Republic and managed large cross-cultural studies.

Alastair Jamieson-Lane is a PhD candidate in applied mathematics at the University of British Columbia. He currently uses mathematics to model biochemical systems and enjoys side projects involving interesting data and collaboration in distant fields.