

Appendix 1: A Cognitive Anthropological Perspective on Cultural Sharing and Variability

CCT offers a concrete formal approach to specify shared cultural knowledge and quantify the extent of such sharing (Romney et al. 1986; Romney et al. 1987), which has allowed field-working anthropologists to fruitfully apply cognitive anthropology in their research. Many prominent figures in psychological, medical, and other forms of anthropology have employed the method, with the original 1986 CCT article eventually becoming one of the most heavily cited journal articles in the discipline as a whole (Atran and Medin 2008; Boster and Johnson 1989; Chavez et al. 2001; Dressler et al. 2015; Garro 2000; Gravlee et al. 2005; Handwerker 2002a; Weller 2007).

With interviews, observations, or more structured methods like free lists, CCT researchers first devise a series of statements that are seen to constitute a cognitive domain of understanding, and then obtain informant judgments about each item in the domain (Romney and Weller 1988). For example, Weller, one of the coauthors of the original 1986 CCT paper, elicited from her Guatemalan informants their views concerning the role of contagion in the etiology of 27 illnesses (Romney 1999; Romney et al. 1986; Weller 1984).

Using the statistical methods originally described in that 1986 paper, traditional CCT researchers measure each individual's so-called cultural competence, or the extent to which that person knows the group's shared model of understanding about the domain in question, which is analogous to linguistic competence (Keesing 1974). As operationalized within CCT, competence values range from 0 to 1 and measure the probability that an informant knows the culturally correct answer to any item within the domain. So, for example, if an informant in Weller's study

had a competence of 0.7, that would mean she had a 70% chance of giving the culturally correct (i.e., shared) answer to any given question about contagion in illness.

CCT algorithms also use competence scores to estimate the group's so-called answer key, or shared responses about a given domain of knowledge, taking into account that the answers of individuals with higher competence scores should weigh more heavily in determining the group's shared or cultural truth.¹

A last technical feature is that the statistical criterion most commonly used to assess the fit of the model to the data is the ratio of the first to second eigenvalues from the factor analysis typically used to estimate the competence scores (but see Lacy and Snodgrass [2016] for an alternative fit measure). In this context, roughly speaking, this ratio measures the extent to which a single cultural group accounts for the patterns of agreement among informants. By convention, an eigenvalue ratio of 3:1, along with the absence of negative competence scores, has been a traditional rule of thumb indicating adequate fit of a single common cultural model. In the analysis of Weller's Guatemalan illness and contagion data, the results correspond to an eigenvalue ratio of 8.6:1, indicating potential consensus in this domain of knowledge (Romney et al. 1986).

Beyond these particular methodological techniques, CCT shares with other anthropological work a cognitive or knowledge approach to culture (D'Andrade 1995; Goodenough et al. 1996).² In this context, an individual's cultural competence score is a measure of how much she or he knows about a particular cognitive domain, relative to other members of the group. As presented thus far, CCT emphasizes shared and thus *consensual* knowledge. But as many scholars have noted, the technique also maps cultural *variability* and lack of sharing, as shown by the variation in individuals' competence scores. For this reason, CCT practitioners

often link their research practice to so-called *distributive* or *epidemiological* views of culture (Rodseth 2008; Sperber 1985). That is, a community's shared knowledge and values are not uniform or homogeneous (Dressler et al. 2015), but differentially distributed across individuals and even across places and particular contexts in the world (Hutchins 1995; Strauss and Quinn 1998). CCT thus clearly intends to capture diversity in sharing, but always relative to a presumed underlying single consensus.

Focused more directly on cultural variation is research in the CCT residual analysis tradition, which explores how members of certain subgroups (identified by variables like gender, age, etc.) differentially adhere to the culturally correct answer key (Atran and Medin 2008; Dressler et al. 2015; Medin et al. 2006; Ross 2004). Residual analysis relies on several techniques, some well-established, others still evolving (Dressler et al. 2015). What matters here is that these so-called residual analysis methods all imply, like the original CCT model, a single answer key underlying responses in the domain, despite the attention to deviations from that single key.

Finally, in all these approaches, we would emphasize that the idea of consensus and dissensus depends in part on the level of analytical abstraction. In standard CCT cases, shared culture is understood to at least partially structure how members model a particular cognitive domain, acknowledged in the analytical practice of fixing CCT questions and the range of possible responses. We might think of this in terms of a common discourse, worldview, or set of underlying values, which at least in part sets the terms of debate and disagreement within a particular place. For example, we see something similar to common underlying values related to freedom and individual responsibility in contemporary mainstream U.S. culture in relation to political thinking among Republicans and Democrats (Lakoff 2002). In those political contexts,

common values and frames of reference set the stage for vehement argument and disagreement. In that U.S. case, we can thus talk both about some level of shared agreement and patterns of difference. Depending on our analytical frame of reference, we might try to establish a more abstract and generally shared cultural frame of reference (and answer key), from which each individual, whether Republican or Democrat, might deviate somewhat. So-called residual analyses might further clarify how such patterns of deviation reflect subgroup differences. Or, we might try and identify more radically different forms of disagreement, which manifest themselves in even unique subcultural Republican and Democratic answer keys. The latter is closer to the approach we sketch in this article, a member of a family of multiple consensus truth methods, whose contours we trace in the article's main body.

Appendix 2: Parametric Bootstrap for Comparing Two Known Groups

To implement the parametric bootstrap test, we follow conventional hypothesis testing procedure, provisionally assuming the null hypothesis as true (i.e., that the data truly derive from a one-group model, so that any difference of TCS_{diff} from 0 merely reflects chance). Under this null assumption, we treat the observed competence values from the conventional one group CCT analysis as the best available estimates of the true distribution of informant competences under a one group model. A computer program then uses these assumed competence values, along with the random response features entailed in the standard CCT guessing assumption described in Romney et al. (1986), to simulate sample datasets of responses for all items and all informants. This simulates the randomness that would arise from informants' responding from a single key according to the assumptions of the original CCT response model (i.e., the conditions of the null hypothesis). This simulation of responses is repeated, say, 1,000 times.

For each such set of simulated data for the sample, a conventional one group CCT solution is calculated. Then, the sample is divided randomly into two groups equal in size to the two known groups, and separate one-group CCT solutions are calculated for each one. Thus, the simulated data incorporates two kinds of randomness, that due to the random component of response assumed in the CCT model as well as that introduced by individuals being randomly shuffled into groups. Each such set of random data represents one instance of what would occur if the null hypothesis were true, that is, that any difference between observed groups could arise randomly. These analyses of each simulated sample set yield, respectively, a value of TCS_{1key} and TCS_{2key} , hence a value of TCS_{diff} for each simulated sample. The collected values for this statistic randomly generated across the 1,000 (or more) repetitions of simulated samples approximate the distribution of the TCS_{diff} values that would occur were the null hypothesis true.

The proportion of simulated TCS_{diff} values exceeding the TCS_{diff} obtained for the observed two *known* groups in the sample gives a formal p -value for the hypothesis test of a one-group ($TCS_{diff} = 0$) versus the alternative of a two-group ($TCS_{diff} > 0$) model. If the observed value of TCS_{diff} is large relative to what occurs by chance simulation, the bootstrap p -value will be small, providing evidence against the one-group (single key) null hypothesis and in favor of the alternative that the fit of a two-group model in the sample is genuinely better, beyond what might arise from the random aspect of informant responses under one group CCT assumptions.

To investigate the performance of this procedure for comparing two known subgroups, we generated and analyzed simulated data to examine: (1) how often this test procedure might mistakenly identify a sample's response data as arising from two keys when in fact it came from one key (i.e., Type I error); and (2) how often the test would correctly identify a sample whose responses derived from two different keys (Power). For the first question, we performed two experiments as follows: A synthetic data set of $N = 30$ simulated individuals was established with randomly generated competence scores (from a beta distribution) with a mean competence score of 0.67. Using these competence values, and the accompanying key, and the standard CCT model assumptions (Romney et al. 1986), we created 1,000 simulated sets of sample response data with 20 questions having 3 responses per question. Each sample was generated using the null hypothesis condition of a single answer key. To simulate the values of TCS_{diff} that can arise from chance rather than the presence of two answer keys, the synthetic informants in each simulated sample were randomly assigned to one of two groups, $N_1 = 20$ and $N_2 = 10$. For each such data set, we applied the bootstrap test (with a significance level of 5%) to see how often it identified the TCS_{diff} between these two random groups as larger than expected by chance (i.e., erroneously contradicting the null hypothesis and therefore falsely indicating that the data

derived from two different keys [groups]). Across all 1,000 simulated data sets, our bootstrap procedure incorrectly rejected the (true) null hypothesis of single group composition 44 times, slightly smaller than expected from our target test size of $\alpha = 0.05$. We also conducted the same experiment but using an initial data set with a relative low level of competence (mean of 0.4), which resulted in 50 Type I errors in 1,000 repetitions. Thus, our test procedure maintained the nominal test size (Type I error), and did not excessively support an incorrect conclusion of two groups/keys.

In a second set of simulations, we investigated the power of this bootstrap test (i.e., its ability to correctly identify samples with responses that indeed arose from two different answer keys). To investigate this, we proceeded as before, with randomly generated data for 1,000 samples each with data for $N = 30$ informants, 20 questions, and three responses per question. However, in this case, data for 10 of the informants in each sample derived from one answer key, while 20 derived from another. The two answer keys differed for two out of the 20 questions and were the same on the other 18. For each group of simulated informants, the mean competence with respect to its own group's key was 0.67. For each sample, we used our bootstrap procedure with a fixed α -level of 0.05 to test the one key/group hypothesis. In 979 out of 1,000 samples, the procedure *correctly* indicated that this hypothesis should be rejected, meaning 98% of the time, it correctly identified data containing two groups of informants responding according to two (only slightly) different answer keys, a quite high value for power by conventional standards.

A comprehensive investigation of the Type I error rate and power of our proposed procedure would entail investigating the role of different sample sizes, different levels, and distributions of competence within each group, varying amounts of difference between the two answer keys, and so on, a task too complex to take on here, so we offer the preceding limited

examination of Type I error and power as proof of concept rather than a definitive demonstration of the properties of our test procedure.

Appendix 3. Maximum Likelihood Estimation of Competence Scores for a One-group Cultural Consensus Model

To conduct conventional cultural consensus analyses, which our known and unknown groups procedures require, we have created our own procedure, available within the more comprehensive computer package to do the various analyses described here. (This is available free as an *R* package [Meyer et al. 2013]). In this appendix, we briefly sketch the maximum likelihood procedure we have used to estimate competence scores for one-group CCT models under the standard assumptions of Romney et al. (1986). While one might choose to obtain these estimates by other means (e. g., by factor analysis), we find the maximum likelihood approach convenient and well-grounded.

For such one group models, we assume that N respondents independently answer K questions, with L options per question. Informants answer questions correctly if they know the answer, and randomly guess one of the L response options with equal probability if they do not know. Let d_i be the competence of the i^{th} informant—i.e., the probability that she or he knows the answer to a randomly selected question—and let π_{ij} be the probability that the answers of informants i and j match on any given question. Finally, let M_{ij} be the number of questions out of K on which the answers of informants i and j are observed to agree or match. Given that observed data, we use the method of maximum likelihood to estimate the competencies. The Bernoulli distribution is used, giving the likelihood of the observed matches first as a function of the π_{ij} :

$$\prod_{i=1}^{N-1} \prod_{j=i+1}^N \pi_{ij}^{M_{ij}} (1 - \pi_{ij})^{K - M_{ij}}$$

from which the log-likelihood is:

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[M_{ij} \ln(\pi_{ij}) + (K - M_{ij}) \ln(1 - \pi_{ij}) \right]$$

The probability π_{ij} that the answers match can be written as a function of the unknown competencies. Based on the standard CCT assumptions per Romney et al. (1986), we compute:

$$\begin{aligned} \pi_{ij} &= d_i d_j + d_i(1 - d_j)1/L + d_j(1 - d_i)1/L + (1 - d_i)(1 - d_j)1/L \\ &= d_i d_j (1 - 1/L) + 1/L \end{aligned}$$

Thus, the log-likelihood function to be maximized over \mathbf{d} is:

$$\psi(\mathbf{d}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[M_{ij} \ln \{ d_i d_j (1 - 1/L) + 1/L \} + (K - M_{ij}) \ln(1 - d_i d_j) \right]$$

This function is concave in \mathbf{d} , and the i^{th} term of the gradient is:

$$\frac{\partial \psi(\mathbf{d})}{\partial d_i} = \sum_{j \neq i}^N \left[M_{ij} \frac{d_j (1 - 1/L)}{d_i d_j (1 - 1/L) + 1/L} - (K - M_{ij}) \frac{d_j}{1 - d_i d_j} \right]$$

This means we can use a gradient-based numerical optimization algorithm, and it will converge to the maximum likelihood estimates of the d_i quickly and reliably. We have implemented this *MLE* approach as a function in the *R* package described above, and have used it whenever estimates for the one-group CCT are needed in the analyses presented here.

Appendix 4: Genetic Algorithm Approach for Identifying Multiple Cultural Groups.

$TCS_{diff-max}$

We start from the assumption that latent groups of informants within the sample might belong to different groups (subcultures) with different answer keys, and thus respond to questions according to their own group's key and their competence with respect to it. We assume that individuals respond using their own group's answer key, according to the standard CCT model assumptions, per Appendix 3. On this basis, we developed a method to determine an optimal separation of individuals into such latent groups. "Optimal" here means that the fit of the model, as measured by the total competence score (TCS) for the identified groups is at a maximum. For simplicity of description here, we consider only the base case in which the analysis posits the existence of two latent groups. Considering all 2^N possible assignments of informants into two groups, and picking the one with the maximum TCS score, would take much too long for even 40 informants, with 2^{40} being approximately equal to one trillion. Consequently, we recommend and have used a so-called genetic algorithm to search among the possible assignments.

A genetic algorithm is a well-established optimization method that uses ideas from evolutionary theory to search a set (in this case, the collection of all possible assignments of N individuals to groups) for its "best" member. Here, each possible assignment of informants can be represented as a string of N ones and twos for two groups. In the language of genetic algorithms, that string is a "phenotype" consisting of N "genes." The method starts with the random generation of a population (of size P , say) of phenotypes. Each phenotype has a "fitness" value, for which we use the TCS computed for the two groups defined by the string of genes. Thus, some members of the population have greater fitness than others.

The population is allowed to evolve using a simulated reproduction step and a mutation step in every “generation.” For the reproduction step, a scheme is required to ensure the more fit phenotypes are more likely to reproduce, but with sufficient randomization to explore the entire set. In our implementation, a “father” is randomly chosen from those strings with fitness in the top 10% of the population. The “mother” is chosen at random from the entire population of strings of genes representing group assignments. The “child” is a random combination of genes from both parents. In every generation, the top 10% of the population is retained, and the others are replaced by children, with fitness again defined by the sum of the estimated competences obtained by calculating the competences for each of the groups using the one-group cultural consensus routine, and adding them together. For the mutation step, a number of phenotypes is randomly selected, and for each of these, a gene is randomly selected and flipped from 1 to 2 or from 2 to 1. Then the fitness for each new generation is re-computed, using the TCS as the criterion. In this way, the overall fitness of the population increases over the generations, and eventually the most-fit phenotype occurs, and the population fills up with phenotypes having the identical genes. The algorithm stops when the top quarter of the population has identical genes. In the current case, this means that the division of the sample corresponds to a maximum value of TCS_{2key} . When compared to the observed sample value of TCS_{1key} , this gives a maximum value of $TCS_{diff-max} = TCS_{2key} - TCS_{1key}$.

Because of the random nature of genetic algorithms, there is no guarantee that the optimal grouping is found. However, if the algorithm is restarted from with a new random collection of phenotypes, and the same answer is obtained, this is considered good evidence that the true maximum has been found.

Appendix 5: Hypothesis Test for Comparing Two Unknown (Latent) Groups

Implementing this test requires approximating the null distribution of $TCS_{diff-max}$, as described in Appendix 4, so as to see what possible values of $TCS_{diff-max}$ might arise from a process in which informants actually respond based on a single key, but which by the random features of CCT generates sample data that to some extent shows evidence of two latent groups (and thus also two answer keys)

The procedure is as follows:

- 1) Apply a conventional one group/key CCT to the sample, obtaining and retaining the observed value for TCS_{1key} , the competence values, and the answer key.
- 2) Apply the genetic algorithm of Appendix 4 to the observed data to find the division of the sample into two groups that gives a maximum value of TCS_{diff} for the observed data, i.e. $TCS_{diff-max}$ as described in Appendix 4. (This summarizes the strength of evidence in the observed sample for the existence of some latent group division, and hence serves as the test statistic.)
- 3) Proceeding under the null assumption that individuals' responses derive from a single key, use the competence values and key obtained from the one-group CCT procedure in step 1) to generate simulated response data for (say) 1,000 samples, at the observed size, using the usual CCT response assumptions.
- 4) For each such simulated sample, obtain a value of $TCS_{diff-max}$, per steps 1 and 2. The collected set of 1,000 such values simulate the sampling distribution of $TCS_{diff-max}$ under the null hypothesis.
- 5) Compare the observed value of $TCS_{diff-max}$ to the simulated sampling distribution, obtaining a p -value for the hypothesis test as the proportion of simulated samples whose

$TCS_{\text{diff-max}}$ exceeds that found from applying the genetic algorithm procedure to the observed sample data.

This estimation of the null distribution of the test statistic follows classical bootstrap ideas. Although the number of possible subgroups is quite large, and the true distribution impossible to calculate, this method provides an accurate Type I Error with reasonably good power. More precision could be attained by increasing the number of bootstrap simulations, at the cost of computation time

Appendix 6: CCMK Applied to Identify Unknown Groups in Simulated Data

We illustrate the unknown groups method first with a small set ($N=20$, 10 questions, 3 response categories) of artificial data shown in Table 1. A conventional one group CCT solution indicated a relatively modest fit, with $TCS_{1key} = 12.4$, and a mean competence score of 0.62 across the 20 individuals. The second eigenvalue ratio from using factor analysis to obtain the

[Table 1 about here]

competences was 2.6, which compared to the conventional criterion of 3.0 raises some question about the fit of a one group model. The reader will notice, in the “Total Sample” column of Table 2 that several individuals (e.g., ID =1) had low competence relative to the key for the total sample, which suggests that it did not capture the culture of all individuals. We then applied the CCMK genetic algorithm procedure to find the “best” division of the sample into two groups with the result being that 15 individuals fit best into one group, and 5 into a second. The “Group” labels (1, 2) in the table indicate those groups. The two latent group keys differed on items 7-10. The sum of the competence scores for this two group solution was $TCS_{2key} = 15.0$, for a mean competence of 0.75, about 21% higher than for the one group solution. The eigenvalue ratios for a cultural consensus model estimated for each of the two groups separately were much larger, 81.5 and 8.6, vs. the original 2.6. Recognizing that such an apparently large improvement in fit of a two group over a one group model might nevertheless arise from the random component of individuals' responses under a one-group model, we conducted the unknown groups bootstrap test just described above. It gave a p -value of 0.001, strong evidence against the null hypothesis that the data derive from a single key.

Notes

1. Sampling of informants within CCT is beyond the scope of the current article, but is an area worthy of further discussion, and treated in some detail in recent work by Handwerker and his collaborators (Handwerker, 2002b; Handwerker, Hatcherson, & Herbert, 1997; Handwerker & Wozniak, 1997).
2. We work within a cognitive anthropology tradition that treats cultures as systems of shared knowledge, meanings, and values, which shape group members' thoughts and orient them toward action (D'Andrade, 1995; Kroeber & Kluckhohn, 1952; Kuper, 2000; Ross, 2004; Strauss & Quinn, 1994, 1998). Within such theories, a key idea is that consensus on a domain of knowledge arises through explicit group member-to-member teaching and thus transmission, as well as through more subtle and indirect forms of social learning (such as through imitation of experts) that signals a common cultural heritage (D'Andrade, 1995).

References

- Atran, S., and D. L. Medin. 2008. *The native mind and the cultural construction of nature*. Cambridge, MA: MIT Press.
- Boster, J. S., and J. C. Johnson. 1989. Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist* 91:866–89.
- Chavez, L. R., J. M. McMullin, S. L. Mishra, and F. A. Hubbell. 2001. Beliefs Matter: Cultural beliefs and the use of cervical cancer-screening tests. *American Anthropologist* 103:1114–29.
- D’Andrade, R. G. 1995. *The development of cognitive anthropology*. Cambridge: Cambridge University Press.
- Dressler, W. W., M. C. Balieiro, and J. E. Dos Santos. 2015. Finding culture change in the second factor stability and change in cultural consensus and residual agreement. *Field Methods* 27:22–38.
- Garro, L. C. 2000. Remembering what one knows and the construction of the past: A comparison of cultural consensus theory and cultural schema theory. *Ethos* 28:275–319.
- Goodenough, W., D. Levinson, and M. Ember. 1996. *Encyclopedia of cultural anthropology*. New York: Henry Holt.
- Gravlee, C. C., W. W. Dressler, and H. R. Bernard. 2005. Skin color, social classification, and blood pressure in southeastern Puerto Rico. *American Journal of Public Health* :2191–97.
- Handwerker, W. P. 2002a. *Quick ethnography: A guide to rapid multi-method research*. Walnut Creek, CA: AltaMira.

- Handwerker, W. P. 2002b. The construct validity of cultures: Cultural diversity, culture theory, and a method for ethnography. *American Anthropologist* 104:106–22.
- Handwerker, W. P., J. Hatcherson, and J. Herbert. 1997. Sampling guidelines for cultural data. *Field Methods* 9:7–9.
- Handwerker, W. P., and D. F. Wozniak. 1997. Sampling strategies for the collection of cultural data: An extension of Boas's answer to Galton's problem. *Current Anthropology* 38:69–75.
- Hutchins, E. 1995. *Cognition in the wild*. Cambridge, MA: MIT press.
- Keesing, R. M. 1974. Theories of culture. *Annual Review of Anthropology* 3:73–97.
- Kroeber, A. L., and C. Kluckhohn. 1952. *Culture: A critical review of concepts and definitions*. Papers. Peabody Museum of Archaeology & Ethnology. Cambridge, MA: Harvard University Press.
- Kuper, A. 2000. *Culture: The anthropologists' account*. Cambridge, MA: Harvard University Press.
- Lakoff, G. 2002. *Moral politics: How liberals and conservatives think*. Chicago: University of Chicago Press.
- Medin, D. L., N. O. Ross, and D. G. Cox. 2006. *Culture and resource conflict: Why meanings matter*. New York: Russell Sage Foundation Publications.
- Meyer, M. C., J. G. Snodgrass, and M. Lacy. 2013. SubCultCon: Maximum-likelihood cultural consensus analysis with sub-cultures (Version 1.0). <https://cran.r-project.org/web/packages/SubCultCon/index.html> (accessed April 19, 2018).
- Rodseth, L. 2008. Distributive models of culture: A Sapirian alternative to essentialism. *American Anthropologist* 100:55–69.

- Romney, A. K. 1999. Culture consensus as a statistical model. *Current Anthropology* 40:S103–S115.
- Romney, A. K., W. H. Batchelder, and S. C. Weller. 1987. Recent applications of cultural consensus theory. *American Behavioral Scientist* 31:163–77.
- Romney, A. K., and S. C. Weller. 1988. *Systematic data collection*, vol. 10. Newbury Park, CA: Sage.
- Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88:313–38.
- Ross, N. 2004. *Culture and cognition: Implications for theory and method*. Thousand Oaks, CA: Sage.
- Sperber, D. 1985. Anthropology and psychology: Towards an epidemiology of representations. *Man* 20:73–89.
- Strauss, C., and N. Quinn. 1994. A cognitive/cultural anthropology. In *Assessing cultural anthropology*, edited by R. Borofsky, 284–300. New York: McGraw-Hill.
- Strauss, C., and N. Quinn. 1998. *A cognitive theory of cultural meaning*, Publications of the Society for Psychological Anthropology, vol. 9. Cambridge: Cambridge University Press.
- Weller, S. C. 1984. Cross-cultural concepts of illness: Variation and validation. *American Anthropologist* 86:341–51.
- Weller, S. C. 2007. Cultural consensus theory: Applications and frequently asked questions. *Field Methods* 19:339–68.