Cultural Consensus Theory: Aggregating Responses in a Finite Interval

William Batchelder*
Alex Strashny A. Kimball Romney
Institute For Mathematical Behavioral Sciences
University of California Irvine

*Department of Cognitive Sciences
whbatche@uci.edu

Work supported by AFOSR Grant to Batchelder and NSF grant to Romney and Batchelder Co-PIs
AGENDA

I. What is Cultural Consensus Theory?
II. A CCT Model for Continuous Responses
   A. Axioms for the Model
   B. Bayesian Statistical Inference
   C. A Cross-Cultural Example
I. WHAT IS CULTURAL CONSENSUS THEORY?

- Cultural Consensus Theory (CCT) is an approach to information pooling (aggregation, data fusion).

- One has access to ‘informants’ who may share
  ✓ ‘cultural’ knowledge or beliefs unknown to the researcher.

  ✓ The researcher can construct questionnaire items but does not know the ‘consensus answers’, if any, that best represent the shared knowledge.

  ✓ Also the researcher does not know the ‘cultural competence’ and ‘response bias’ of each informant nor the ‘cultural saliency’ (difficulty) of each question.
ORIGIN OF CCT

- CCT was created by A. Kimball Romney, and myself in the 1980s. Email me for a reference list (whbatche@uci.edu)

- Romney, Weller, & Batchelder (1986), “Culture as Consensus: A Theory of Culture and Informant Accuracy” first introduced CCT to the Anthropology community. It is the most cited article in American Anthropologist, the flagship journal of the American Anthropological Society, since its origin in 1888 (cf. web of science)

- CCT has become a major methodological tool in social, cultural and medical anthropology, e.g. Weller, 2007, Field Methods. Also it has been used in other areas of the social and cognitive sciences, e.g. psychometrics, social networks.
Some CCT APPLICATION AREAS

- Ethnographic studies in social and cultural anthropology, for example determining folk medical or religious beliefs
- Determining beliefs of a deviant group
- Inferring events that happened from eyewitness reports
- Discovering social relationships in a social network
- Pooling Judges’ scorings of contests
- Determining the syntax of an exotic language
- Aggregating probabilistic forecasts
II. A. The Continuous Response Model

- Examples of data situations for responses in a finite interval:
  * Degree of confidence in the truth of a proposition.
  * Probability estimates of the likelihood of events.
  * Rating the similarity of two semantic terms.

- There have been models by others that can be applied to ordinal or continuous response data without prior knowledge of the answers, e.g. Johnson and Albert, 1999, *Ordinal Data Analysis*, Springer.

- Our approach is different in that it attempts to provide a cognitively plausible model of the informant.
The Data Set Up

- We model responses that fall in the unit interval $I = [0, 1]$; other finite intervals can be transformed linearly to $[0, 1]$.

- We follow an approach used in models for psychophysical judgments by postulating that there are latent random variables $Y_{ik} \in (0, 1)$, that capture the internal representation produced in informant $i$ by item $k$, $1 \leq i \leq N, 1 \leq k \leq M$.

- The manifest responses $X = (X_{ik})_{N \times M}$ are a function of the corresponding latent random variables $Y_{ik}$; however, the model includes a bias function that can distort the $Y_{ik}$ due to each respondent’s characteristics (response biases) in using the scale.
Informal Setup

The cultural truth

The latent representations

The observable responses

Estimation of the cultural truth and other model parameters

\[ Z_k \in (0,1) \quad \text{(A Model parameter)} \]

\[ Y_{ik} = Z_k + e_{ik} \quad \text{(Error Enters)} \]

\[ X_{ik} = h(Y_{ik}; b_i) \quad \text{(Response Bias Enters)} \]

\[ \hat{Z}_k \in (0,1) \quad \text{(Bayesian Estimation)} \]
Core Axioms for Model

Axiom 1. (Common Truth). There is a fixed answer key,
\[ Z = \langle z_k \rangle_{1 \times M} , \text{ where } z_k \in I \text{ for } k = 1, 2, \ldots, M. \]

Axiom 2. (Latent Representations). The latent representational random variables are given by
\[ Y_{ik} = z_k + e_{ik} , \text{ where the } e_{ik} \text{ are continuous type random variables representing measurement error, with } E(e_{ik}) = 0 \]
\[ -z_k < e_{ik} < 1 - z_k , \text{ and variances } \sigma_{ik}^2 > 0 . \]
Core Axioms Continued

Axiom 3. (Inhomogeneous Competence). There is a function $\phi$ from $I$ into the positive reals given by $\phi(z) = z(1 - z)$ and competencies $D = \langle D_i \rangle_{1 \times N}$, with $D_i > 0$, such that

$$\sigma_{ik}^2 = \phi(z_k)/(1 + D_i)$$

Axiom 4. (Conditionally Independent Errors). The $e_{ik}$ are conditionally independent given the parameters $Z = \langle z_k \rangle_{1 \times M}$ and $D = \langle D_i \rangle_{1 \times N}$.

- These core axioms can be augmented in several ways to specify the error distribution and the response bias in how the informant uses the scale to generate responses.
Properties of the Core Axioms

- It is possible to estimate the parameters, the \( (Z_k)_{1xM} \) and \( (D_i)_{1xN} \), by a method of moments procedure based on the correlations between every pair of informants over items

\[
r_{ij} = \sum_{k=1}^{M} (X_{ik} - \bar{X}_i)(X_{ik} - \bar{X}_j) / S_i \cdot S_j
\]

- The Model satisfies a version of Spearman’s famous law of tetrads: \( \forall 1 \leq i \neq j \leq N, \rho_{ij} = \rho_i Z \rho_j Z \)

Where \( \rho_i Z \) is the correlation between informant i’s responses and the latent cultural truth. This leads to a testable implication of the model- one dimensionality of

\[
R = (r_{ij})_{N \times N}
\]
Introducing an Error Distribution

- The beta distribution is a convenient distribution for the error.

\[ \forall x \in (0,1), \ f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} (1-x)^{\beta-1}; \alpha, \beta > 0 \]

\[ E(X) = \frac{\alpha}{(\alpha + \beta)} \quad Var(X) = \frac{\mu(1-\mu)}{(\alpha + \beta + 1)} \]

- Core Axioms \[ \Rightarrow \ E(Y_{ik}) = z_k \quad Var(Y_{ik}) = z_k (1-z_k)/(1+D_i) \]

**Axiom 5. (Beta Distributed Errors).** The \( Y_{ik} \) have beta distributions with parameters given by

\[ \alpha_{ik} = z_k \cdot D_i \quad \beta_{ik} = (1-z_k) \cdot D_i \]
Modeling Response Bias

- The informant must translate their internal representation into a manifest scale value $X_{ik} = h(Y_{ik})$. We bias over and under use of the scale by a one parameter function subject to: $h(0;b)= 0$, $h(1/2;b)= 0.5$, $h(1;b)= 1$, and for $0 < y < 1$, $h(y;b)= 1-h(1-y;b)$, $0<b<1$.

- We formulated $h(y;b)$ as a spline of two quadratic Bezier curves subject to the above conditions (details omitted)

$$ h(y,b) = \begin{cases} 
\frac{\left( b - 2y \cdot b - 1 + y + \sqrt{1 - 2b + b^2 - 2y + 4y \cdot b} \right) \cdot (2b - 1)^{-1}}{y} & \text{if } 0 \leq b \neq 0.5 \\
\frac{\left( 3b - 1 + y - 2y \cdot b - \sqrt{-1 + 2b + b^2 + 2y - 4y \cdot b} \right) \cdot (2b - 1)^{-1}}{y} & \text{if } 0.5 < y < 1 \land b \neq 0.5 \\
0 & \text{if } b = 0.5
\end{cases} $$

OK Sorry email me for the details
Figure 1 shows three Bézier bias function for the points $b=0$, $b=0.5$, $b=1$.

**Axiom 6. (Bézier Bias).** The manifest response random variables, $X_{ik}$, are given by $X_{ik} = h(Y_{ik}; b_i)$ for parameters $b_i \in (0, 1)$, $i = 1, \ldots, N$, $k = 1, \ldots, M$. 

![Diagram showing Bézier bias functions with points at b=0, 0.5, and 1]
II. Bayesian Statistical Inference

- Parameters: \( Z = (Z_k)_{1 \times M} , D = (D_i)_{1 \times N} , b = (b_i)_{1 \times N} \)

- Posterior distribution

\[
p[Z, D, b \mid X = (x_{ik})] = \frac{L[(x_{ik}) \mid Z, D, b] \pi(Z, D, b)}{\int_{ZD} L[(x_{ik}) \mid Z, D, b] \pi(Z, D, b) \, dZ \, dD \, db}
\]

- Usual approach sets up a Markov Chain Monte Carlo Sampler using the fact that

\[
p(\bullet) \propto L(\bullet) \pi(\bullet)
\]

- Probability of the data-a constant but with M+2N parameters can’t get
Metropolis Hasting Algorithm

- We developed an MCMC sampler based on the Metropolis-Hasting (M-H) algorithm to estimate a Bayesian fixed effects version of the model following suggestions from Carlin & Louis (2000, *Bayes and Empirical Bayes Methods for data Analysis*).

- The prior assumed the parameters were independent with uniform distributions on the $Z_k$ and $b_i$ and diffuse exponential distributions on the $D_i$. - Email me for details.

- Work with Monte Carlo data generated from the model convinced us that the MCMC sampler recovered the parameters to a satisfactory degree.
III. A Cross-Cultural Example

- The purpose of the study was to compare the perceptions of the similarity of emotion terms between English and Japanese monolingual speakers.

- The dataset consisted of $N=65$ respondents and $M=105$ items. The respondents fell into two groups: 33 monolingual English speakers interviewed in the United States and 32 monolingual Japanese speakers interviewed in Japan.

- The items consisted of all 105 pairs from 15 selected terms to describe emotions, e.g. bored (tsumaranai), fear (osoroshii), and shame (hazukashii)- For original study see Romney, Moore, Rusch, 1997, P. Natl. Acad. Sci. USA.
Experimental Procedure

- Each respondent was required to rate each of the 105 pairs of emotion terms for perceived similarity on a five point likert scale, with 1 most dissimilar and 5 most similar.

- Yes, this isn’t continuous data. Simulation suggested it was ‘OK’, but we are working on a CCT model for a Likert scale and also to collecting similarity data on a continuous scale.

- Both cultures had similar distributions of competence and bias. Bias was in the direction of scale expansion, but with lots of individual differences.
# Cross-cultural Comparisons

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Culture</th>
<th>Mean</th>
<th>Median</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_i$</td>
<td>English</td>
<td>9.26</td>
<td>9.24</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>9.33</td>
<td>9.28</td>
<td>3.76</td>
</tr>
<tr>
<td>$b_i$</td>
<td>English</td>
<td>0.28</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>0.34</td>
<td>0.28</td>
<td>0.21</td>
</tr>
</tbody>
</table>

![Distribution of Competence](image1)

![Distribution of Bias](image2)
Multidimensional Scaling

- The means of the estimated $z_k$ for each pair were submitted to a metric multidimensional scaling program using methods to compare the two cultures (see Kumbasar, Romney, Batchelder, 1997, *Am. J. Sociol.*).

- The results were surprising to ‘Whorfians.’ While there were some significant differences, e.g. ‘shame,’ the cultures viewed the emotion terms very similarly both in terms of their clusters and the closeness of their multidimensional placements.

- The original study used correspondence analysis on the raw similarity responses, our methods are designed to smooth out bias and differentially weight informants - results were sharper.
A representation of the mean placements of the 15 emotion terms for the English (star) and Japanese (circle) groups in the first two dimensions based of a principle component metric multidimensional scaling of the consensus answer key parameters of the CCT model.
Conclusion

- There is ongoing work to: (1) to develop Bayesian hierarchical inference for the informant parameters, (2) to incorporate item difficulty parameters, (3) to investigate other types of response bias.

- The model is being applied to some real continuous data involving similarity in several semantic systems as well as the perception of difficulty of indoor rock climbing routes.

- We think it may be a valuable tool for cross-cultural comparisons as well as pooling forecasts.