10/24/2011

Cultural Consensus Theory:

Detecting Experts and their Shared Knowledge

William H. Batchelder

Professor Cognitive Sciences

UCI

Co-workers: A. Kimball Romney, Kalin Agrawal , Royce Anders, Ece Batchelder, John Boyd, George Karabatsos, Zita Oravecz, Alex Strashny, Susan Weller.

Support is gratefully acknowledged from grants from NSF, the Air Force Office of Scientific Research, Army Research Office, and IARPA

Papers available by request: whbatche@uci.edu

AGENDA

I. Detecting Experts Without Performance Information

II. What is Cultural Consensus Theory (CCT)?

III. The General Condorcet Model (GCM)

- A. Axioms for the Model
- B. Properties and Inference
- C. Applications of the GCM
- D. Further Developments of the GCM

IV. Other CCT Models

I. Detecting Experts Without Performance Information

- A researcher often locates experts in some domain of ability by seeing who performs well on a relevant ability test.

- Psychometric test theory provides numerous formal ways to assess expertise using especially designed statistical models.

- However test theory requires that a researcher knows the answers to the questions on the ability test.

- What if a researcher knows enough to ask relevant questions about some domain of expertise but lacks any knowledge about the ‘correct answers’ to the questions?

A Thought Exercise

- Suppose you are a grader in a graduate quantum mechanics course. You have not been to class and know nothing about the subject.

- The Professor has given a true/false exam to *N* students, but alas you have lost the answer key. What might you do?

- The General Condorcet Model to follow is an approach that estimates the correct answers to the questions and the relative knowledge of the students, both endogenously. It beats the majority rule see:

✓ Batchelder, W.H., Romney, A.K. (1988) Test theory without an answer key. *Psychometrika,* **53**, 71-92.

✓ Karabatsos, G.A., Batchelder, W.H. (2003). Markov chain estimation theory methods for test theory without an answer key, *Psychometrika*, **68**, 373-389.

Test Theory Without an Answer Key

- Since the mid 1980s, myself and colleagues have been developing an area called ‘Cultural Consensus Theory’ (CCT).

- In essence CCT is designed to simultaneously assess the relative ability of people who share expertise in some domain as well as the nature of the expertise that they share.

- CCT does this by using formal cognitive models to analyze response patterns to test data, where both the ability level of respondents and the ‘consensus answers’ to the questions are specified as parameters in the model.

II. What is Cultural Consensus Theory?

- CCT is an approach to information pooling (aggregation, data fusion) developed by A. Kimball Romney, myself, and colleagues (see reference sheet).

- CCT is has become a popular methodology in the social sciences especially social and cultural anthropology.

- CCT is suited for cases where several informants share knowledge or beliefs that are unknown apriori to the researcher.

- An essential feature of CCT is that there is no requirement that the shared knowledge corresponds to a ‘ground truth.’

Some CCT Application Areas

- Ethnographic studies in social and cultural anthropology, e.g. determining folk medical beliefs
- Determining consensus beliefs of a deviant group
- Determining perceived or actual relationships in a social network
- Determining the syntax of an exotic language
- Aggregating forecasts from different experts, e.g. weather or elections.
- Pooling judges rankings of contestants in sports competition, e.g. ballroom dancing, boxing, gymnastics
- Determining what actually happened from eyewitness reports.

Types of Questions and Models

- ❖ The questions represent knowledge or beliefs shared by a group rather than individual preferences. For example:

    Do the Hoosiers like basketball?-- OK

    Do you like basketball? -- Not OK

- ❖ We develop questions in various formats, e.g. true/false; multiple choice, Likert, matching, ranking, continuous scale.

- ❖ For each question format, we formulate cognitive models of how a respondent answers the questions. The models are often like models in psychometric test theory except the ‘culturally correct answers’ are parameters in the model instead of being known apriori.

The Nature of the Data

- We select a questionnaire format, generate *M* items pertaining to the unknown shared knowledge, and collect responses to each question from each of *N* informants.

- The data is represented in an *NxM* response profile matrix

$$\mathbf{X} = (X_{ik})_{NxM}$$

  where X_{ik} is the response of informant *i* to question *k*.

- Some missing data is OK, but **X** should not be too sparse.

Goals of CCT

- We want to use the information in $\mathbf{X} = (X_{ik})_{NxM}$ to decide if the statistical model we use is valid in assuming that the group shares common cultural knowledge.

IF 'YES' TO:

1. Estimate the consensus answers and cultural saliency (difficulty) of each question
2. Estimate the level of expertise and response bias characteristics of each informant

IF 'NO' TO:

1. Determine if the informants cluster into several cultural groups, each with its own answer key.
2. If so, detect the groups either with covariates or by using a finite mixture version of the statistical model.

III. The General Condorcet Model

- Each of *N* informants answer “yes” or “no” to each of *M* items.
- Response Profile Data-- $\mathbf{X} = (X_{ik})_{NxM}$

$$X_{ik} = \begin{cases} 1 \text{ if informant } i \text{ answers "yes" to item } k \\ 0 \text{ if informant } i \text{ answers "no" to item } k \end{cases}$$

- Answer Key-- $\mathbf{Z} = (Z_k)_{1xM}$

$$Z_k = \begin{cases} 1 \text{ correct answer to item } k \text{ is "yes"} \\ 0 \text{ if correct answer to item } k \text{ is "no"} \end{cases}$$

- Performance Profile Data-- $\mathbf{Y} = (Y_{ik})_{NxM}$

$$Y_{ik} = \begin{cases} 1 \text{ if informant } i \text{ correctly answers item } k \\ 0 \text{ if informant } i \text{ answers incorrectly to item } k \end{cases}$$

Relationship Among Random Variables

- In test theory, the answers **Z** are known, **X** is observed from test takers, and the performance **Y** is determined.

- In CCT, **X** is observed, and **Z** and **Y** are latent (unobserved).

- **OBSERVATION 1:** Given any two of **X**, **Z**, **Y**, we can determine the third by $\forall 1 \le i \le N, 1 \le k \le M$

$$X_{ik} = Y_{ik} Z_k + (1 - Y_{ik})(1 - Z_k)$$

$$Z_{ik} = X_{ik} Y_{ik} + (1 - X_{ik})(1 - Y_{ik})$$

$$Y_{ik} = X_{ik} Z_k + (1 - X_{ik})(1 - Z_k)$$

GCM Parameters

- The GCM specifies parameters for the 'culturally correct' or consensus answers to the questions, namely , $\mathbf{Z} = (Z_k)_{1xM}$, with space $Z_k \in \{0,1\}$

- In addition the model specifies hit rate and false alarm rate parameters (see **Axiom 3**), for each informant denoted, respectively, by $\mathbf{H} = (H_i)_{1xN}, \mathbf{F} = (F_i)_{1xN}$, $0 < F_i \leq H_i < 1$

- As we will see, the GCM is a prototype of various yes/no signal detection models, except the hit and false alarm rates are latent instead of observed.

III.A. Axioms for the GCM

AXIOM 1. (Single Culture). There is a single answer key $\mathbf{Z} = (Z_k)_{1xM}$ applicable to all informants.

AXIOM 2. (Conditional Independence). The response profile matrix satisfies conditional independence given by

$$\Pr[\mathbf{X} = (x_{ik})_{NxM} \,|\, \mathbf{Z}, \mathbf{H}, \mathbf{F}] = \prod_{i=1}^{N} \prod_{k=1}^{M} \Pr(X_{ik} = x_{ik} \,|\, Z_k, H_i, F_i)$$

for all possible realizations $(x_{ik})_{NxM}$

AXIOM 3. (Marginal Probabilities). $\forall 1 \le i \le N, \forall 1 \le k \le M,$

$$\Pr(X_{ik} = 1 \,|\, Z_k, H_i, F_i) = Z_k H_i + (1 - Z_k) F_i$$

Specifying Hits and False Alarms

- As with all signal detection type models, hits and false alarm probabilities are transformed to parameters that separately tap signal detectability (expertise) and response bias.

- The GCM uses the double high threshold model (2HTM)

- **Axiom 4. (2HTM).** There are informant expertise parameters **D**= $(D_i)_{1xN}, 0 < D_i < 1$, and informant response bias parameters **G**= $(g_i)_{1xN}, 0 < g_i < 1$, such that

$$H_i = D_i + (1 - D_i)g_i, \; F_i = (1 - D_i)g_i$$

III.B. Properties and Inference

- ❖ For Bayesian inference, we need the likelihood function

OBSERVATION 2: The likelihood function from Axioms 1-4 is given by

$$L[\mathbf{Z},\mathbf{D},\mathbf{G} \mid (X_{ik})] = \prod_{i=1}^{N}\prod_{k=1}^{M}[(D_i + (1-D_i)g_i]^{Z_k X_{ik}}[(1-D_i)(1-g_i)]^{Z_k(1-X_{ik})}$$
$$[(1-D_i)g_i]^{(1-Z_k)X_{ik}}[D_i + (1-D_i)(1-g_i)]^{(1-Z_k)(1-X_{ik})}$$

- ❖ Note that each of the four settings of the terms in the exponents retains only one of the four terms in the product.

A Spearman Property of the GCM

- Let K be a random variable with space the first M positive integers and probability distribution

$$\Pr(\mathrm{K}=k)=\begin{cases}1/M \text{ if } k\in\{1,\ldots,M\}\\ 0 \text{ otherwise}\end{cases}$$

- Define the correlation between pairs of informants over items

$$\rho(X_{i\mathrm{K}},X_{j\mathrm{K}})=\frac{Cov(X_{i\mathrm{K}},X_{j\mathrm{K}})}{\sqrt{Var(X_{i\mathrm{K}})Var(X_{j\mathrm{K}})}} \quad 1\le i\ne j\le N$$

- **OBSERVATION 3:** Assume the GCM in Axioms 1-3.

$$\forall 1\le i\ne j\le N,\ \rho(X_{i\mathrm{K}},X_{j\mathrm{K}})=\rho(X_{i\mathrm{K}},Z_{\mathrm{K}})\rho(X_{j\mathrm{K}},Z_{\mathrm{K}})$$

The Spearman Property is Testable

- Observation 3 says that the correlation between two informants over items is the product of their correlations with the unknown truth.

- This is Spearman's tetrad law signaling a single factor.

$$\forall \text{ distinct } 1 \le i, j, k, l \le N; \ \rho_{ij}\rho_{kl} = \rho_{il}\rho_{kj}$$

- It implies the testable condition that the off diagonal terms of the observed informant by informant correlation matrix approximately satisfies

$$\mathbf{R} = \left(r_{ij}\right)_{NxN} \approx \left(a_i\right)_{Nx1}\left(a_i\right)_{Nx1}^T$$

where the $a_i > 0$ are estimates of the latent ρ_{iZ} .

Why the Model Beats the Majority Rule

- If the informant parameters, **D** and **G,** are known, then the answer key can be estimated item by item by examining the odds ratio

$$LR_k = \frac{\Pr(Z_k = 1 \mid <X_{ik}>_{i=1}^{N})}{\Pr(Z_k = 0 \mid <X_{ik}>_{i=1}^{N})}$$

- **OBSERVATION 4 :** Bayes theorem, and assuming $g_i \equiv 1/2$

Log odds

$$\hat{Z}_k = 1 \text{ iff } \ln LR_k = \sum_{i=1}^{N} (2X_{ik} - 1)\ln\left(\frac{1+D_i}{1-D_i}\right) + \ln\frac{\tau}{(1-\tau)} \geq 0$$

where τ is the prior probability that 'yes' is the culturally correct answer.

Statistical Inference

- The original method of estimation we developed in the 1980s (see references) was a sort of hybrid of a method of moments to get the D_i , and a Bayes approach to get the Z_k, much like the last slide (Batchelder & Romney, 1988, *Psychometrika*).

- The method was clever but statistically inefficient. Unfortunately it caught on and has been the usual one for several hundred applications in Anthropological studies.

- The main objective of CCT is to estimate the consensus answers. For the GCM they are discrete so combinatorial optimization has been used, e.g. simulated annealing (Batchelder, Kumbasar, & Boyd, 1997,*J. Math. Sociol*.).

Bayesian Inference for the GCM

- More recently Bayesian inference with MCMC methods has been developed (Karabatsos & Batchelder, 2003, *Psychometrika*). This approach numerically estimates the posterior distribution of parameters given data.

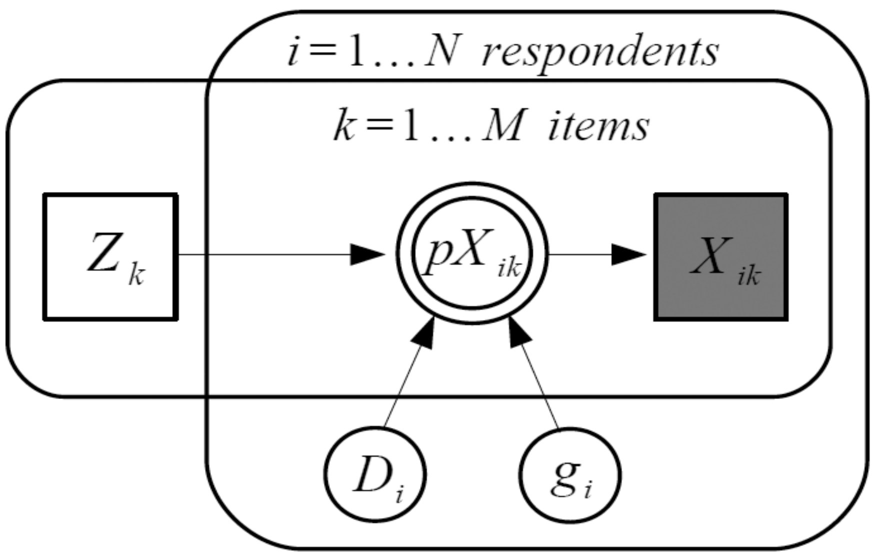
$$g(\Theta|\mathbf{D}) \propto \pi(\Theta)L(\Theta|\mathbf{D})$$

- $L(.|.)$ is the likelihood function of the model and $\pi(.)$ is the prior selected by the researcher.

- We developed our own MCMC sampler in S, and the program is accessible from George Karabatsos' website at University of Illinois Chicago Circle.

DAG Representation of Models

- Lately we have been using JAGS and WinBUGS. All my students are up on this now, and we are developing a general user friendly software program for a variety of CCT models, thanks to a grant facilitated by Jeff Johnson and the ARO.



$$Z_k \sim Bernoulli(.5)$$

$$D_i \sim Uniform(0,\ 1)$$

$$g_i \sim Uniform(0,\ 1)$$

$$pX_{ik} = z_k D_i + (1 - D_i) g_i$$

$$X_{ik} \sim Bernoulli(pX_{ik})$$

III.C. Applications of the GCM

- There have been over 200 applications of the GCM, mostly in cultural anthropology. Usually the goal is to estimate the consensus answers to the questions, but also estimating levels of expertise is also important.

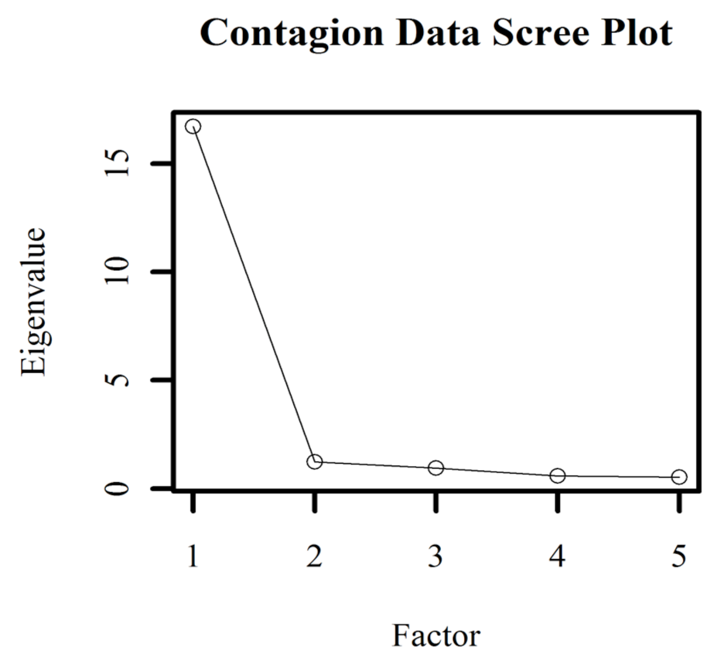
- For example, in a study by Dr. Susan Weller, she asked 24 women from an urban Guatemala community for each of 27 diseases (e.g. arthritis, colic, …) whether or not they were 'contagious' and whether they needed a 'hot remedy' or a 'cold remedy.'

- The contagious data but not the hot/cold data showed strong cultural consensus when analyzed with a restricted version of the GCM model. We reanalyze both sets of data with JAGs.

Some Statistics from the Posterior Distribution

Contagion Data	Z	D	G
Mean	0.52	0.74	0.55
SD		0.10	0.20

The scree plot comes from a minimum residual factor analysis of the informant by informant correlation matrix discussed earlier it exhibits a strong one factor solution



Information from Posterior Distributions

Ignore right side with LTM

		GCM Inference Results						LTM Inference Results					
	Node	***Mean***	***SD***	***2.5%***	***50.0%***	***97.5%***	***Rhat***	***Mean***	***SD***	***2.5%***	***50%***	***97.5%***	***Rh***
	1	0.94	0.06	0.77	0.95	1.00	1.00	33.74	11.14	10.71	35.27	49.32	1.0
	2	0.93	0.06	0.77	0.95	1.00	1.01	33.70	11.31	9.64	35.49	49.47	1.0
	3	0.86	0.09	0.63	0.88	0.98	1.00	31.94	11.46	9.56	33.01	49.09	1.0
	.	.	.	.	.	.	.	.	.	.	.	.	.
D_i / E_i	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.
	22	0.60	0.14	0.29	0.61	0.84	1.00	16.60	10.38	4.48	13.91	43.84	1.0
	23	0.53	0.15	0.21	0.54	0.78	1.00	6.45	2.52	4.04	5.67	13.15	1.0
	24	0.53	0.15	0.23	0.55	0.78	1.00	7.95	4.06	4.09	6.74	18.99	1.0
	1	0.50	0.29	0.03	0.50	0.98	1.00	0.46	0.08	0.32	0.46	0.62	1.0
	2	0.51	0.29	0.02	0.52	0.98	1.00	0.46	0.08	0.31	0.46	0.61	1.0
	3	0.69	0.23	0.18	0.73	0.99	1.00	0.42	0.07	0.28	0.42	0.57	1.0
	.	.	.	.	.	.	.	.	.	.	.	.	.
g_i / b_i	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.
	22	0.29	0.15	0.05	0.27	0.61	1.00	0.58	0.08	0.43	0.58	0.74	1.0
	23	0.38	0.15	0.12	0.38	0.68	1.00	0.53	0.09	0.36	0.53	0.71	1.0
	24	0.25	0.14	0.04	0.23	0.57	1.00	0.63	0.09	0.45	0.63	0.79	1.0
dev		402.2	9.13	386.1	401.6	421.6	1.00	304.6	10.0	287.0	303.9	326.3	1.0
	1	1.00	0.00	1.00	1.00	1.00	1.00	0.87	0.09	0.67	0.89	1.00	1.0
	2	1.00	0.00	1.00	1.00	1.00	1.00	0.87	0.09	0.67	0.89	0.99	1.0
	3	1.00	0.00	1.00	1.00	1.00	1.00	0.87	0.09	0.67	0.89	0.99	1.0
	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.
z_k / T_k	.	.	.	.	.	.	.	.	.	.	.	.	.
	25	0.00	0.00	0.00	0.00	0.00	1.00	0.10	0.07	0.00	0.09	0.26	1.0
	26	0.00	0.00	0.00	0.00	0.00	1.00	0.10	0.07	0.00	0.09	0.26	1.0
	27	0.00	0.00	0.00	0.00	0.00	1.00	0.10	0.07	0.00	0.09	0.25	1.0

Sample Output From WinBUGS

- As part of a cross cultural study by Dr. Ece Batchelder and myself, N=74 So.Cal. healthy elderly Hispanics answered M=157 T/F about causes, symptoms, and treatment options about AD.

Key	Mean	SD	Comp	Mean	SD	.025	Mdn	.975
Z(1)	.09	.28	D(1)	.68	.06	.55	.68	.79
Z(2)	1.00	.00	D(2)	.73	.05	.63	.74	.83
Z(3)	0.00	.02	…					
…	…		Bias	Mean	SD	.025	Mdn	.975
			g(1)	.36	.09	.20	.35	.54
			g(2)	.75	.06	.64	.75	.86
			…					

III. D. Further Developments of the GCM

- In several papers we have generalized the GCM to handle:
  1. Differential cultural saliency (difficulty) of the items. Karabatsos & Batchelder(2003)
  2. Incorporating covariates of the ability parameters. Zita Oravecz and Batchelder (in preparation)
  3. Tests for multiple answer keys along with an associated finite mixture model for multiple answer keys.Batchelder & Romney, 1989; Anders & Batchelder, In preparation).
  4. Applications of the GCM to determine ties in a social network subject to side constraints on the answer key. Batchelder, Kumbasar, Boyd (1997), Kalin and Batchelder (In preparation).

Re 1: THE TWO-HIGH THRESHOLD RASCH MODEL

- We model the knowledge parameters with a special parameterization of the Rasch Item Response Theory (IRT) Model.

$$D_{ik} = \frac{\theta_i(1-\alpha_k)}{\theta_i(1-\alpha_k)+\alpha_k(1-\theta_i)}$$

$$0 < D_{ik}, \mathrm{g_i}, \theta_i, \alpha_k < 1; i = 1,\ldots,N; k = 1,\ldots,M$$

This version uses the Rasch model on a latent variable D_{ik} . The consequence is that

$$\log it(D_{ik}) = \log it(\theta_i) - \log it(\alpha_k)$$

ability Item difficulty

Re:2 Adding Covariates

❖ In practice one may know a vector of *C* covariates on each informant

$$(V_{ic})_{NxC}$$

❖ Then one can regress the latent ability parameters on the covariates with an appropriate link function *h*

$$h(\sum_{c=1}^{C} \beta_c V_{ic} + \varepsilon_i) = D_i$$

Where ε_i is an appropriate error term. This is running in JAGs and a paper is in preparation (Oravecz and Batchelder).

Re 3: Testing for Multiple Cultures

- In case the evidence is strong that the one-culture GCM fails to hold, there are two possibilities:

1. The model is valid except there are two or more cultures, each with its own answer key.
2. The data are 'hash', there is no consensus signal in the data.

- Batchelder and Romney (1989) suggested tests concerning the factoring of the correlation matrix $\left(\rho_{ij}\right)_{NxN}$ that can decide between #1 and #2 above.
- Royce Anders and I have extended the GCM to allow more than one answer key using techniques from FMMs. We have a running Bayesian inference using JAGs.

Re 4: Application to Graph Data with Constraints on the Answer Key.

- ❖ Batchelder, Kumbasar, and Boyd (1997) applied the GCM to digraph data concerning friendship relations in a hardware engineering business.

- ❖ Lately Kalin Agarwal and I have been developing applications to graphs where the answer key, but not the informants, are restricted to satisfy side axioms like balance or an equivalence relation.

- ❖ To handle the restrictions of the side conditions, special, combinatorially complex samplers have been developed.

IV. Other CCT Models

- So far there are published CCT models for multiple choice, ranking, matching, and items requiring a continuous response in a finite interval.

- Models have been and are being developed for ranking items and items on a Likert scale.

- Email me for more details.

SELECTED REFERENCES

- Batchelder, W.H., Kumbasar, E., and Boyd, J.P. (1997). Consensus analysis of three-way social network data. *Journal of Mathematical Sociology*, **22**, 29-58.
- Batchelder, W.H. and Romney, A.K.(1986). The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman and G. Owen (eds.) *Information Pooling and group decision Making: Proceedings of the Second University of California Irvine Conference on Political Economy* (pp.103-112).Greenwich, Conn.: JAI Press .
- Batchelder, W.H. and Romney, A.K. (1988). Test theory without an answer key. *Psychometrika*, **53**, 71-92.
- Batchelder, W.H. and Romney, A.K. (1989). New results in test theory without an answer key. In E.E. Roskam (Ed.) *Mathematical Psychology in Progress* (pp.229-248).Heidelberg: Springer-Verlag.
- Batchelder, W.H. (2009) .Cultural consensus theory: Aggregating expert judgments about ties in a social network. In H. Liu, J. Salerno, and M.J. Young (Eds.). Social Computing, Behavioral Modeling , and Prediction 2009. New York: Springer, 2009, pp. 24-32.
- Batchelder, W.H. Cognitive Psychometrics: Using Multinomial Processing Tree Models as Measurement Tools. In S. E. Embretson (Ed.). Measuring Psychological Constructs: Advances in Model Based measurement. Washington DC: American Psychological Association Books, 2009, Ch.4.

• Batchelder, W.H., Strashny, A., and Romney, A.K. Cultural Consensus Theory: Aggregating Continuous Responses in a Finite Interval. In S.-K. Chai, J.J. Salerno, and P.L. Mabry (Eds.). Social Computing, Behavioral Modeling , and Prediction 2010. New York: Springer, 2010, pp. 98-107.
• Karabatsos, G., and Batchelder, W.H. (2003). Markov chain estimation theory methods for test theory without an answer key. *Psychometrika*, **68**, 373-389.
•Kumbasar, E., Romney, A.K., and Batchelder, W.H. (1994). Systematic biases in social perception. *American Journal of Sociology*, **100**, 477-505.
•Romney, A.K., and Batchelder, W.H. (1999). Cultural consensus theory. In R. Wilson and F. Keil (eds.).The MIT Encyclopedia of the Cognitive Sciences. Cambridge, MA: The MIT Press.
•Romney, A.K., Batchelder, W.H., Weller, S.C. (1987). Recent applications of cultural consensus theory *American Behavioral Sciences*. **31**, 129-141.
•Romney, A.K., Weller, S.A., and Batchelder, W.H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, **88**, 313-338.
•Weller, S.A. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods*, **19**, 339-368.