An Extended Cultural Consensus Theory Model to Account for Cognitive Processes in Decision Making in Social Surveys

Zita Oravecz, Katherine Faust and William H. Batchelder

Sociological Methodology 2014 44: 185 originally published online 22 May 2014
DOI: 10.1177/0081175014529767

Sociological Methodology
2014, Vol. 44(1) 185–228

DOI: 10.1177/0081175014529767
http://sm.sagepub.com

3

AN EXTENDED CULTURAL CONSENSUS THEORY MODEL TO ACCOUNT FOR COGNITIVE PROCESSES IN DECISION MAKING IN SOCIAL SURVEYS

Zita Oravecz*
Katherine Faust*
William H. Batchelder*

Abstract

In recent decades, cultural consensus theory (CCT) models have been extensively applied across research domains to explore shared cultural knowledge and beliefs. These models are parameterized in terms of person-specific cognitive parameters such as abilities and guessing biases as well as item difficulties. Although psychometric test theory is also formalized in terms of abilities and item difficulties, a quality that clearly sets CCT models apart from other test theory models is their specification to operate on data in which the answer key is latent. In doing so, CCT models specify the answer key as parameters of the model, and also involved with this specification are procedures to verify the integrity of the answer key that is estimated. In this article, the authors develop the following methods to propagate the application of these CCT models in the field of social surveys: (1) by extending the

*University of California, Irvine, USA

Corresponding Author:
Zita Oravecz, University of California, Irvine, Department of Cognitive Sciences, 3213 Social and Behavioral Sciences Gateway Building, Irvine, CA 92697-5100, USA
Email: zoravecz@uci.edu

underlying cognitive model to be able to account for uncertainty in decision making ("don't know" responses), (2) by allowing covariate information to be entered in the analysis, and (3) by deriving statistical inference in the hierarchical Bayesian framework. The proposed model is fit to data describing knowledge on science and on aging to demonstrate the novel findings that can be achieved by the approach.

Keywords

social survey, cultural consensus theory, Bayesian hierarchical model, don't know (DK) response

1. INTRODUCTION

Cultural consensus theory (CCT) arose in cultural anthropology, in which ethnographers are often faced with the task of identifying cultural knowledge in a particular population (Romney, Weller, and Batchelder 1986; Romney and Batchelder 1999). In a typical situation, the cultural knowledge is not known to the ethnographer ahead of time, so a central goal of CCT is to provide a systematic method for inferring the consensus answers by pooling the responses to a set of questions pertaining to a particular domain of knowledge (Romney et al. 1986; Romney and Batchelder 1999). CCT relies on formal cognitive and measurement models to examine respondents' shared cultural knowledge or beliefs. Typically, a set of respondents answer various questions concerning a specific topic. Consensus models describe the cultural consensus by estimating a shared answer key while accounting for the knowledge levels and response styles of the respondents.

A central tenet of CCT is that cultural understandings arise both through direct experience and through learning from other members of a culture. Cultural knowledge is thus a product of social context and process. As a consequence, knowledge and understanding of various content domains come to be more or less shared among members of the same culture. This being the case, we expect there to be some degree of within-culture consensus with respect to domain-specific knowledge. In sociology, this contextual view of knowledge is paralleled to some extent in the social construction of knowledge tradition of Berger and Luckmann (1966), who described how a substantial amount of knowledge is the product of society and of social processes. CCT formalizes this theoretical insight and offers quantitative measures for the level of cultural consensus in a group, for the consensus answer, and for cognitive characteristics of the members of the group. In this way, CCT models are able to directly measure cultural knowledge without a prior hypothesis as to the culturally agreed upon answers. In contrast, traditional methods for summarizing responses to knowledge-based questions usually focus on analyzing performance relative to expert established correct answers.

In this article, we introduce a CCT-based model to the field of survey research to contribute to understanding of consensus knowledge in general and to enable investigation of sources of individual differences in decision making involved in responding to social survey questions. CCT is potentially appropriate for any domain in which one expects socially constructed knowledge or a belief system shared among members of a population, society, or culture. Examples of CCT applications include folk beliefs of the causes and treatments of diseases (Weller 1984a, 1984b; Hruschka et al. 2008; Baer et al. 2003; Weller et al. 1999), properties and uses of plants (Hopkins 2011), ecological relations among species (Atran et al. 1999; Medin et al. 2006), judgment of personality traits in social networks (Batchelder, Kumbasar, and Boyd 1997; Agrawal and Batchelder 2012), national consciousness (Yoshino 1989), and extracting truth from eyewitness testimonies (Waubert de Puiseau et al. 2012).

1.1. *Latent Consensus Answer Key*

One of our goals in this article is to demonstrate the applicability of CCT to survey research on knowledge or belief areas. In these types of questionnaires, typically there is a list of questions measuring some knowledge domain (e.g., science, environment, treatment of illnesses), and most likely respondents' answers are scored as correct or incorrect on the basis of an expert established correct answer for each question. In contrast, in our approach, the established correct answers are not used in analyzing the data, but rather a consensus set of answers is estimated by fitting a CCT model, which then allows us to compare these consensus solutions with other sets of answers for the same questions, such as answers resulting from scientific investigations. However, we emphasize that this method as applied here is not aimed at ascertaining what the truth is. For example, in case of science knowledge items, we do not claim to obtain scientifically valid answers with the CCT approach.

Rather, we derive what the social consensus is on items—and along with that, we test whether there is consensus at all—regardless of whether that consensus is actually true. We offer a method to elicit shared opinions from groups of respondents and intend to treat those as culturally held opinions, not as natural or objective facts. Such socially constructed consensus knowledge is an interesting topic of research in its own right, as it is manifestly present in society, in a form that is relatively consistent, persistent, and measurable. Although traditional methods, such as item response theory (IRT)–based models (De Boeck and Wilson 2004) can reveal the lack of knowledge, CCT is apt for investigating the potential existence of "counter-knowledge" or "biased knowledge." For example, CCT models have been applied to medical knowledge and beliefs (e.g., see Weller et al. 2012 or, for a summary, Weller 2007) to investigate how the recovered answer key departs from scientific medical knowledge.

1.2. *Uncertainty in Decision Making: "Don't Know" Answers*

In this article, we focus on survey items with dichotomous response alternatives (i.e., true or false). Because people are not always willing to commit to responding either "true" or "false," surveys often permit a "don't know" (DK) response to accommodate this possibility. In that case, responding "don't know" indicates a legitimate state of a respondent's knowledge. Providing a DK option has been widely discussed. One stream of research on DK responses concerns whether there are sociodemographic or cultural differences in the tendency to respond "don't know" or to offer no opinion (Francis and Busch 1975; Bauer 1996; Krosnick et al. 2002). Another stream considers the consequences of including DK responses: on one hand, whether the elimination of DK responses reduces the effects of differences in guessing biases (Mondak and Davis 2001; Mondak and Canache 2004) and gives a more valid representation of knowledge or, on the other hand, whether the inclusion of DK responses allows one to distinguish uninformed from misinformed responses (Sturgis, Allum, and Smith 2008) and reduces measurement error introduced by random guessing (Courtenay and Weidemann 1985; Luskin and Bullock 2011). The source of DK answers can be attributed to characteristics of the questions (e.g., the question is ambiguous or illogical, the question wording is unconventional) and characteristics of the respondents (e.g., motivation to answer correctly, socioeconomic status). Therefore, it has been pointed out by several researchers (Clogg 1982, 1984; Mondak and Canache 2004) that DK responses should not be simply considered as incorrect responses, as being incorrect translates to false knowledge rather than to uncertainty in knowledge. It has also been shown that DK responses should not be coded as missing data (no responses) either. Because DK answers are not random, treating them as missing data leads to sample selection bias (Feick 1989; Liao 1995). As an alternative, association models (Clogg 1982, 1984), latent class models (Clogg 1984; Feick 1989), and correction through sample selection bias models (Liao 1995) have been proposed.

We present here an extended version of the general CCT model to account for uncertainty in decision making as measured by DK responses. The most popular and widely used CCT model, called the general Condorcet model (GCM; described later), is used to analyze dichotomous responses (true or false) to questions pertaining to a specific domain of culturally shared knowledge. Several research studies applying the GCM have resorted to augmenting DK responses in ad hoc ways, for example, by coin flipping, as shown in the summary of the problem in Weller (2007). Although CCT research in ethnographic applications generally finds a small number of DK responses, social surveys can have fairly large numbers of DK responses, which means that such unprincipled reassignment of DK responses would be questionable. Moreover, from a cognitive and consensus modeling perspective, it is important to distinguish between an uninformed response (''don't know'') and a misinformed response (an ''incorrect'' answer), because these are permissible knowledge states. As an alternative to data augmentation, an extended Condorcet model (ECM) can be specified to allow for DK responses, which can then be directly modeled. With the ECM, we can take into account the possibility that when respondents do not know the answer, some are more prone to guessing, whereas others would rather choose the DK option. Moreover, guessing bias is also assumed, which may skew the average response toward agreement or rejection, meaning that guessing in the absence of relevant knowledge is not a fully random process.

Being able to distinguish between uninformed and misinformed responses is important both for appropriately representing knowledge states and for studying knowledge gaps in a society. When responses are scored only as correct or incorrect (with DK responses treated as incorrect), respondents who are more willing to guess can score higher just by chance. For example, it has been shown that women and men differ in their willingness to guess, and a method that scores answers only as correct or incorrect mistakes the difference in performance scores between genders as difference in knowledge level (Mondak and Canache 2004; Lizotte and Sidman 2009). As will be shown, in the proposed cognitive model–based framework, when a DK response is included, it is possible to separate differences in knowledge from differences related to response style.

1.3. *CCT-based Modeling Approach*

We propose the hierarchical ECM (HECM) as a formal approach to model socially constructed knowledge as represented in social surveys. In this framework, we treat person- and item-specific parameters as random effects (De Boeck and Wilson 2004). The resulting hierarchical model allows taking into account interindividual and interitem differences, while pooling information among subjects and items to decrease uncertainty in the parameter estimates. This framework also offers a straightforward way to enter explanatory variables (e.g., gender, age) in the model. To summarize, several novel findings can be achieved by analyzing survey data with the HECM: (1) Consensus knowledge on survey items can be estimated; (2) DK responses can be accounted for by a person-specific response-style parameter; (3) differences in guessing bias can be separated from differences in knowledge level; (4) the general population can be described in terms of HECM parameters such as knowledge level, guessing bias, and willingness to guess; and (5) individual differences in these HECM parameters can be quantified and possibly explained by explanatory variables, including, for example, sociodemographic variables that have long been a focus in research on knowledge gaps. Also, we note that statistical inference and parameter estimation for the HECM are carried out in the Bayesian framework (Gelman et al. 2004; Gill 2007; Kruschke 2011). Parameters in the Bayesian framework have probability distributions, which offers an intuitively appealing way of investigating uncertainty in the estimates.

As a demonstration of our approach, in this article the HECM is applied to two knowledge domains: (1) science knowledge and (2) knowledge on aging. Knowledge survey items are often included in large-scale surveys such as the General Social Survey (GSS). Science knowledge has been included in numerous surveys over the past several decades. Most prominent among these is the National Science Board's ''Science and Engineering Indicators'' (National Science Board 1996, 2006, 2008, 2010, 2012; Miller 1998, 2004, 2011), which includes a standard set of questions about knowledge of scientific concepts and phenomena and is administered as part of the GSS.

With respect to knowledge about aging, a standard questionnaire developed for this purpose is Palmore's Facts on Aging Quiz (FAQ; Palmore 1998). The FAQ analyzed in this article was also part of a larger survey called the 1994 Images of Aging in America (IAA; American Association of Retired Persons 1994). By investigating knowledge on the aging process, potential misconceptions and favorable or unfavorable biases toward the elderly can be explored. However, these application areas are just two illustrations, and other problems investigated by survey research, such as different knowledge domains (e.g., politics, history, popular culture), attitudes, or values, can also be examined by CCT models.

The rest of this article is organized as follows. We first describe the properties of the GCM. Next, this model is extended to be able to incorporate a DK response alternative. The resulting model, the ECM, is formulated as a hierarchical model by assuming that its parameters come from joint population distributions. It is followed by the specifications of statistical inference on the HECM in the Bayesian framework. In the next sections, the model is applied to social survey data from 2010 GSS and 1994 IAA-FAQ. This is followed by posterior predictive model checking and robustness testing. Finally, we summarize the findings and conclude with a discussion of model interpretation and possible extensions.

2. THE GCM

We begin with a short summary of the specifications of the GCM (Romney et al. 1986; Batchelder and Romney 1988), which is a model for dichotomous (true or false) items. A detailed description of the GCM's axioms and properties can be found in Batchelder and Anders (2012). The GCM enables researchers to estimate a consensus answer key from the response pattern of informants who share some knowledge or belief system, while taking into account these respondents' abilities and guessing bias tendencies.

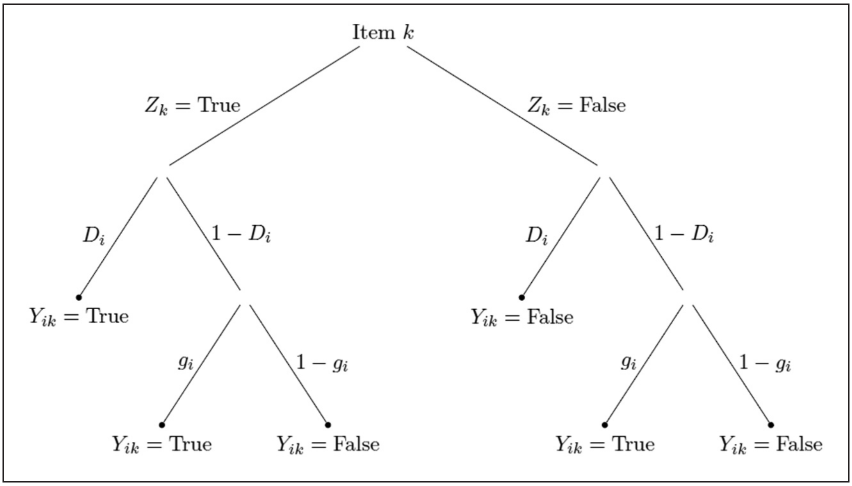

Figure 1. Processing tree of the general Condorcet model for an item k.

The GCM relies on a formal model called the two high-threshold model from signal detection theory (e.g., Macmillan and Creelman 2005). Figure 1 depicts the assumed processing tree of a respondent's decision process. The answer from a single respondent i ($i = 1, \ldots, N$), for item k ($k = 1, \ldots, M$) is denoted as Y_{ik}.

The $N \times M$ data set $\boldsymbol{Y}$ is assumed to have only dichotomous entries:

$$Y_{ik} = \begin{cases} 1 & \text{if } i \text{ responds "True" to item } k \\ 0 & \text{if } i \text{ responds "False" to item } k \end{cases}.$$

The latent consensus answer parameter for item k, denoted as Z_k, is either ''true'' or ''false'' (coded 1 and 0, respectively). It is assumed that there is only one underlying culture behind the data; therefore all respondents share the same answer key $\boldsymbol{Z}$ (see a model check for this assumption later). Respondent i either knows the response for the question with probability D_i or does not know the answer, with probability $1 - D_i$. In the latter case, the respondent guesses ''true'' with probability g_i or ''false'' with probability $1 - g_i$. The model has $2 \times N$ respondent-specific parameters (probability of knowing the answer, $\boldsymbol{D} = [D_i]_{1 \times N}$, and guessing bias, $\boldsymbol{G} = [G_i]_{1 \times N}$) and M item-specific parameters (the latent answer key, $\boldsymbol{Z} = [Z_k]_{1 \times M}$).

The likelihood function for the GCM can be derived by going down the branches of the processing tree and adding up the probabilities and by using the conditional-independence property of the GCM (e.g., Batchelder and Romney 1988):

$$L(\boldsymbol{Z}, \boldsymbol{D}, \boldsymbol{G} | \boldsymbol{Y} = Y_{ik}) = \prod_{i=1}^{N} \prod_{k=1}^{M} [D_i + (1 - D_i)g_i]^{Y_{ik}Z_k} [(1 - D_i)g_i]^{Y_{ik}(1-Z_k)} \\ \times [(1 - D_i)(1 - g_i)]^{(1-Y_{ik})Z_k} [D_i + (1 - D_i)(1 - g_i)]^{(1-Y_{ik})(1-Z_k)}.$$

On the basis of the properties introduced above, it was shown in Batchelder and Anders (2012) that the correlation between two respondents over items equals the product of each respondent's correlation with the answer key, formally for $\forall 1 \leq i \neq l \leq N$:

$$\rho(Y_{iK}, Y_{lK}) = \rho(Y_{iK}, Z_k)\rho(Y_{lK}, Z_k), \tag{1}$$

where K is a random variable that selects a random item index, so that $\forall k, Pr(K = k) = 1/M$.

Equation (1) leads to a consequence that for all distinct respondents denoted as i, l, m, n

$$\rho(Y_{iK}, Y_{lK})\rho(Y_{mK}, Y_{nK}) = \rho(Y_{iK}, Y_{nK})\rho(Y_{mK}, Y_{lK}),$$

which is a form of Spearman's law of tetrads (Spearman 1904). Spearman's application was to correlations between different tests of intelligence, and he interpreted the tetrads law as indicating that intelligence is measured by a single factor. In our case, the tetrad law occurs because there is a single answer key Z_k behind the respondent-by-respondent correlations over the items.

Originally, statistical inference for the GCM was carried out in the classical inferential framework, as discussed in Batchelder and Romney (1988). Later, Karabatsos and Batchelder (2003) derived inferences for different versions of the GCM (including one with item difficulty, as shown later) in the Bayesian framework. Recently, Oravecz, Vandekerckhove, and Batchelder (forthcoming) developed user-friendly graphical user interface–based software that can estimate GCMs in the Bayesian framework. An extension to the GCM incorporating a continuous latent answer key as well as hierarchical Bayesian inference (see later) can be found in Batchelder and Anders (2012).

3. THE ECM

In this section, we extend the GCM to DK responses, this way arriving at the ECM. In the ECM, DK answers are recognized as uncertainty in knowledge.

Consider a simple questionnaire created to collect information on the cultural consensus on a certain topic. The answer from a single respondent i $(i = 1, \ldots, N)$ to item $k (k = 1, \ldots, M)$ is denoted as Y_{ik}. As described above, some respondents might feel unsure about their knowledge on particular items. Therefore on the data level, we allow ''true,'' ''false,'' and ''don't know'' answers. Formally, the data are represented as

$$Y_{ik} = \begin{cases} 1 & \text{if } i \text{ responds “True” to item } k \\ 2 & \text{if } i \text{ responds “False” to item } k \\ 3 & \text{if } i \text{ responds “Don't” know to item } k. \end{cases}$$

Similar to the GCM, in the ECM, it is assumed that each item k has an underlying culturally shared answer Z_k, indicating whether the consensus answer to each item is ''true'' or ''false'' (coded 1 and 0, respectively) according to the cultural consensus within the population.

Figure 2 summarizes the decision tree of respondent i for item k on the basis of the proposed extended cognitive model. The underlying decision process is as follows: Respondent i either knows the consensus answer for item k with probability D_{ik} or does not $(1 - D_{ik})$. In the latter case, the respondent can still decide to guess the correct answer with probability b_i (willingness to guess parameter). If the respondent is not willing to take a guess, it is assumed that he or she marks the ''don't know'' option, with probability $1 - b_i$. In case of guessing, the probability of guessing ''true'' is g_i (the guessing bias parameter).

For example, in the study of folk medical beliefs (e.g., see Weller et al. 1993), it is clear that sometimes respondents answer from a shared consensus at variance with scientific medical knowledge. Knowing the scientifically wrong answer is different from failing to respond with the scientifically correct answer. The latter but not the former can result from guessing or the selection of a DK option.

The probability for each type of answer can be derived by simply going down the branches of the corresponding answers in Figure 2. Formally, the probabilities of the answer categories on the basis of the unsimplified forms derived from the processing tree (see simplified forms later in equation 17) are written as

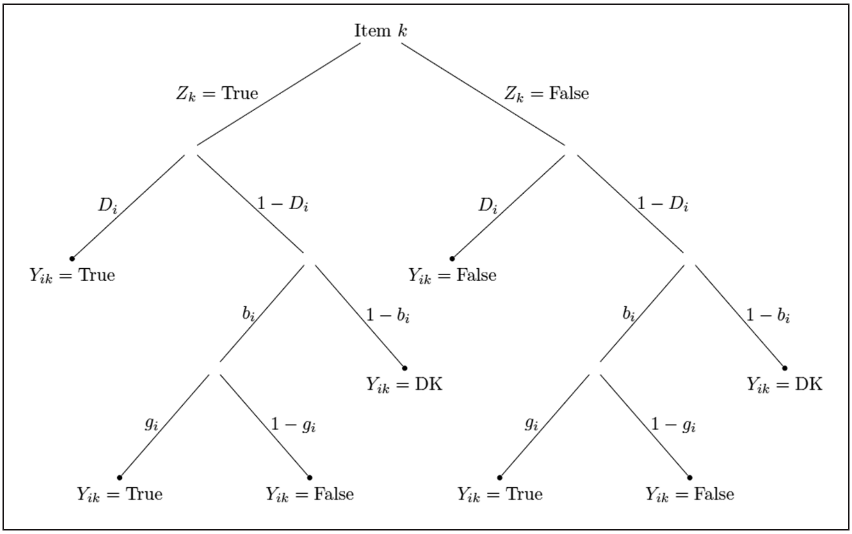

Figure 2. Processing tree of the extended Condorcet model.
Note: DK = don't know.

$$p(Y_{ik} = \text{“True”} = 1) = Z_k[D_{ik} + (1 - D_{ik})b_i g_i] + (1 - Z_k)[(1 - D_{ik})b_i g_i], \tag{2}$$

$$p(Y_{ik} = \text{“False”} = 2) = Z_k[(1 - D_{ik})b_i(1 - g_i)] + (1 - Z_k)[D_{ik} + (1 - D_{ik})b_i(1 - g_i)], \tag{3}$$

and

$$p(Y_{ik} = \text{“Don't know”} = 3) = Z_k[(1 - D_{ik})(1 - b_i)] + (1 - Z_k)[(1 - D_{ik})(1 - b_i)]. \tag{4}$$

The data (Y_{ik}) are then assumed to come from a categorical distribution (denoted as Cat) with a 3×1 probability vector $\boldsymbol{p}_{ik}$ with elements corresponding to the three answer categories:

$$Y_{ik} \sim \text{Cat}(\boldsymbol{p}_{ik}),$$

where the probability vector is $\boldsymbol{p}_{ik} = [p(Y_{ik} = 1)\ p(Y_{ik} = 2)\ p(Y_{ik} = 3)]$, with elements defined in equations (2), (3), and (4).

So far, it has been assumed that all questionnaire items have the same difficulty level. However, Batchelder and Romney (1988) introduced the idea of item heterogeneity in the GCM. Statistical inference for the GCM with item heterogeneity was derived later by Karabatsos and Batchelder (2003). We also believe that in the case of the ECM, the possibility of item heterogeneity should generally be incorporated in the model. The formulation of heterogeneous item difficulty for the ECM departs somewhat from the original proposal by Batchelder and Romney (1988), and it follows the general Rasch model as used in psychometric test theory (or IRT, as in De Boeck and Wilson 2004; Gelman and Hill 2007). In this model, the probability of knowing the correct answer for an item (D_{ik}) is simply a function of the respondent's ability and the item's difficulty level in the following way:

$$D_{ik} = \text{logit}^{-1}(\theta_i - \delta_k), \tag{5}$$

where θ_i is the ability parameter belonging to respondent i, and δ_k denotes the item difficulty level of question k, where $\theta_i, \delta_k \in \mathbb{R}$. Because D_{ik} is a probability parameter, the difference between parameters θ_i and δ_k is transformed back to the unit scale by the inverse-logit function. If their difference is 0, D_{ik} equals 0.5. If the respondent's ability θ_i is greater than the item's difficulty δ_k, then the respondent has a better-than-even probability of knowing the consensus answer to that item; that is, $D_{ik}>0.5$. As can be seen from equation (5), adding the same constant to both parameters would still result in the same D_{ik}. Therefore, to identify the model, some constraints must be introduced. In a nonhierarchical model, one of the item difficulty parameters (δ_k) is typically fixed to a chosen value. In a hierarchical model (discussed in the next section), most often it is the population mean of these parameters that is fixed.

Although the ECM uses the same decomposition for the probability of knowing the correct answer as IRT models, these two approaches have three different goals and assumptions:

1. IRT assumes that the researcher has an a priori answer key for the questions and then scores item difficulty and subject ability accordingly. In contrast, CCT works with response data, not scored performance data. When applying the ECM, the cultural consensus answers are estimated on the basis of patterns of respondents' answers. Therefore, CCT and IRT have fundamentally different goals: IRT is more apt to measure

ignorance in knowledge areas, whereas CCT is able to detect ''counter-knowledge,'' a consensus-based rejection of certain ideas.

2. The proposed ECM accounts for DK responses by a latent cognitive model for the decision-making process, whereas if we were to fit an IRT model, we would have to code the DK responses as incorrect answers (or maybe as missing data).
3. With the ECM, we can study differences in cognitive response style in terms of willingness to guess and guessing bias, and they can be tied to covariates, such as socioeconomic status. In short, the ECM does not simply work on the performance scale, as IRT does, but it models the underlying latent knowledge while taking into account person-specific guessing tendencies and willingness to guess in the decision process.

To summarize, the ECM introduced above has $3\times N$ informant-specific parameters, namely, the ability parameters ($\boldsymbol{\theta} = [\theta_i]_{1\times N}$), the willingness-to-guess parameters ($\boldsymbol{B} = [b_i]_{1\times N}$), and the guessing bias parameters ($\boldsymbol{G} = [G_i]_{1\times N}$). Also, it has $2\times M$ item-specific parameters: the answer key for each item ($\boldsymbol{Z} = [Z_k]_{1\times M}$) and the item-difficulty parameter for each item ($\boldsymbol{\delta} = [\delta_k]_{1\times M}$). The likelihood function can be written as

$$L(\boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{G}, \boldsymbol{B}, \boldsymbol{\delta} | \boldsymbol{Y} = Y_{ik}) = \prod_{i=1}^{N}\prod_{k=1}^{M}(b_i g_i + D_{ik}Z_k - D_{ik}b_i g_i)^{\mathbf{I}_{Y=1}} \times [D_{ik} - D_{ik}Z_k + b_i(D_{ik}-1)(g_i-1)]^{\mathbf{I}_{Y=2}}[(D_{ik}-1)(b_i-1)]^{\mathbf{I}_{Y=3}}, \quad (6)$$

where $\mathbf{I}$ stands for the indicator function that takes the logical value 1 if its subscript is true (and 0 otherwise). Equation (6) is based on simplified forms of equations (2), (3), and (4) that described the probabilities for the answer categories, where D_{ik} is as defined in equation (5). In the next section, population distributions are assigned for each of the parameters.

4. THE HECM

The hierarchical framework allows drawing inferences on the hierarchical (group) level while individual differences are taken into account in a statistically coherent and logical framework (Snijders and Bosker 1999; Raudenbush and Bryk 2002; Gelman and Hill 2007). The hierarchical structure on the model parameters assumes that the parameters of the same type share a certain commonality expressed by their superordinate distributions. When performing CCT analysis hierarchically, we take advantage of the information across respondents and items, which enhances the recovery of the person- and item-level parameters. In the paragraphs that follow, some reasonable superordinate distributions are chosen for each person- and item-specific parameter.

The person-specific ability (θ_i) and the item-specific item difficulty (δ_k) parameters are defined on the real line, which means that independent normal (or Gaussian, denoted with N) distributions can be assigned as their population distribution:

$$\theta_i \sim N(\mu_\theta, \sigma_\theta^2) \tag{7}$$

$$\delta_k \sim N(0, \sigma_\delta^2), \tag{8}$$

where σ_θ^2 and σ_δ^2 represent the variations in these populations, while μ_θ represents the average ability in the population. For model identifiability, the population mean of the item difficulties is fixed to 0, as shown in equation (8).[1] An advantage of identifying the model this way as opposed to fixing one δ_k (or θ_i) to 0 is that all item-specific difficulties (or abilities) are recovered.

In addition to allowing a DK response and accounting for different item difficulties, it is desirable to incorporate explanatory variables (covariates) about the respondents in the model as well. For this purpose, the population mean of the ability parameters can be decomposed into covariate scores and corresponding regression coefficients. More specifically, the covariate information for respondent i on covariate j ($j = 1, \ldots, J$) is denoted as x_{ij}, while regression coefficients are generally denoted with βs with different subscripts. The general guideline to covariate modeling in the HECM is to standardize all covariate scores.

For example, in section 6, ''Applying the HECM to Survey Data,'' we use years of education, gender, age, and religiousness as covariates. Hence, instead of directly modeling a population mean, the mean of equation (7), μ_θ can be replaced by μ_{θ_i}:

$$\mu_{\theta i} = \beta_{\theta 0} + \beta_{\theta,\,\mathrm{gender}} x_{i,\,\mathrm{gender}} + \beta_{\theta,\,\mathrm{age}} x_{i,\,\mathrm{age}} + \beta_{\theta,\,\mathrm{edu}} x_{i,\,\mathrm{edu}} + \beta_{\theta,\,\mathrm{rel}} x_{i,\,\mathrm{rel}}\,. \tag{9}$$

Parameter $\mu_{\theta i}$ here represents the person-specific ability. As mentioned above, because covariate scores $\mathbf{x}$ are standardized, the intercept $\beta_{\theta 0}$ represents the common tendency in the population and can be interpreted as population mean. The rest of equation (9) displays the combination of person-specific covariate scores and regression coefficients, which is the so-called random-effect part of the model (see more on generalized linear models in De Boeck and Wilson 2004). This part allows the explanation of interindividual variation in the person-specific ability estimates ($\mu_{\theta i}$).

To arrive at a more general formulation, all regression coefficients (including an intercept) for the ability parameter θ can be collected in a vector, $\boldsymbol{\beta}_\theta$. Consequently, $\mu_{\theta i}$ can be written as a dot product of the predictors and regression coefficients:

$$\mu_{\theta i} = \mathbf{x}_i^T \boldsymbol{\beta}_\theta . \tag{10}$$

In the event that there is no covariate information available, $\boldsymbol{\beta}_\theta$ in equation (10) reduces to the intercept $\beta_{\theta 0}$, which means that $\beta_{\theta 0} = \mu_\theta$ (for all i).

The other two person-specific parameters (the guessing bias g_i and the willingness to guess b_i) are defined on the interval [0,1], which means that we first apply a logit transformation on these variables to put them on the real line, and then we follow the same hierarchical covariate modeling principle for logit(g_i) and logit(b_i) as for θ_i above:

$$\text{logit}(g_i) \sim N(\mathbf{x}_i^T \boldsymbol{\beta}_g, \sigma_g^2), \tag{11}$$

$$\text{logit}(b_i) \sim N(\mathbf{x}_i^T \boldsymbol{\beta}_b, \sigma_b^2). \tag{12}$$

If the guessing bias and willingness to guess are not made a function of predictors, the population means in equations (11) and (12) reduce to the intercepts $\beta_{g0} = \mu_g$ and $\beta_{b0} = \mu_b$ for all i. The population variance parameters σ_g^2 and σ_b^2 provide information about the level of interindividual variation in terms of guessing bias and willingness to guess, respectively.

The final addition to the ECM model concerns the answer key. Typically, it is assumed that the answer key items are generated hierarchically by a Bernoulli process with a specific hyperprior. The majority of GCMs (e.g., see Batchelder and Romney 1988; Karabatsos and Batchelder 2003) fix the probability parameter of this Bernoulli process to 0.5, this way designating a priori equal chances for every latent answer key parameter to be either ‘‘true’’ or ‘‘false.’’ This constraint could be relaxed, as in some questionnaires, it might be difficult to balance ‘‘true’’ and ‘‘false’’ items because the answer key is unknown a priori. Formally, the answer key items are assumed to come from a Bernoulli (denoted as ‘‘Bern’’) distribution:

$$Z_k \sim \text{Bern}(\pi), \tag{13}$$

where π is the probability of an answer key item being ‘‘true.’’ With this extension, the uncertainty related to the true proportions of ‘‘true’’ and ‘‘false’’ answer key items is directly taken into account when estimating model parameters.

To summarize, the HECM extends the GCM with an extra parameter to take into account willingness to guess, and it also augments it hierarchically to be able to pool information across respondents and items. The next section describes statistical inference for the HECM in the Bayesian framework.

5. BAYESIAN STATISTICAL INFERENCE IN THE HECM

In this section, statistical inference for the model is derived in the Bayesian statistical framework (Gelman et al. 2004; Gill 2007; Kruschke 2011). Although Bayesian methods offer a coherent and principled way of deriving statistical inference, the complexity of the HECM also motivates the choice of this framework. Parameter estimation in the maximum likelihood framework would involve a high-dimensional integration over the several random-effect distributions with no closed-form solutions. An advantage of Bayesian statistical inference is that it focuses on the *posterior density* of the parameters, which can be explored by sampling techniques. In general, the posterior density represents the probability distribution of the parameters given the data, and it can be written as

$$p(\boldsymbol{\gamma}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}),$$

where $\boldsymbol{\gamma}$ denotes the vector of all model parameters, and $\boldsymbol{Y}$ represents the data. The posterior density is proportional to the product of the likelihood of the data given the parameters and the prior distribution of the parameters. The latter represents our prior knowledge about the model parameters. In most studies, noninformative distributions are chosen as priors because we rarely have information on the model parameters beforehand. The more data acquired, the more influential the likelihood function becomes on the posterior, therefore in turn dominating the prior.

When estimating the parameters for the HECM, we explore the marginal conditional posterior distributions of the parameters by taking advantage of Markov chain Monte Carlo methods. This collection of sampling techniques generates samples from these posteriors. For sampling we rely on JAGS, freely available computer software (Plummer 2003). Once the samples meet criteria for convergence and for sufficiently large effective sample size, we can calculate point estimates, posterior standard deviations, posterior credible intervals, and so on.

In the Bayesian hierarchical model described above, the prior information on the person- and item-specific parameters (θ_i, g_i, b_i, δ_k, and Z_k) is contained within their (hyperparametrized) population distributions, as defined in equations (7), (8), (11), (12), and (13). These population distributions have free parameters, namely, the population mean or regression coefficients and variance, which are estimated from the data through their conditional posterior distributions. As an example, the conditional posterior distribution for the variance of the respondent-specific ability parameters can be written as

$$p(\sigma_\theta^2 | \boldsymbol{\theta}_{|\boldsymbol{Y}}, \mu_\theta) \propto p(\boldsymbol{\theta}_{|\boldsymbol{Y}}, \mu_\theta | \sigma_\theta^2) p(\sigma_\theta^2),$$

where the posterior probability of σ_θ^2 depends on the person-specific abilities that are conditional on the data ($\boldsymbol{\theta}_{|\boldsymbol{Y}}$) and on the population mean (μ_θ), while $p(\sigma_\theta^2)$ denotes the prior on that population variance parameter.

In general, we assign diffuse prior distributions for these population parameters—more specifically, a flat normal distribution on the regression parameters (collected in a vector $\boldsymbol{\beta}$) and uniform distribution (denoted as U) on the population standard deviation parameters:

$$\boldsymbol{\beta}_f \sim N_{J+1}(0, 10\boldsymbol{I}_{J+1}) \tag{14}$$

$$\sigma_h \sim U(0.01, 4), \tag{15}$$

where f can be replaced by any of the $F = 3$ model parameters: θ, g, and b; h can be replaced with any of the $H = 4$ model parameters: θ, g, b, and δ; and $\boldsymbol{I}$ stands for the identity matrix. Although the prior on the standard deviation (σ_h) can be considered mildly informative, its range is in correspondence with its role, namely, that it expresses residual variation in the regression equation with standardized scores. Also, in prior sensitivity analysis, we found no remarkable difference when more diffuse reference priors were implemented.

As a prior on the Bernoulli probability for the answer key Z_k, a uniform prior distribution is assigned:

$$\pi \sim U(0, 1)\,. \tag{16}$$

Now the conditional posterior density of all model parameters given the data is derived. For notational convenience, all person-specific parameters are collected into corresponding vectors (i.e., $\boldsymbol{\theta}$, $\boldsymbol{B}$ and $\boldsymbol{G}$). Item difficulties and answer key items are represented as vectors as well (i.e., $\boldsymbol{\delta}$ and $\boldsymbol{Z}$). Then the conditional posterior density can be written as

$$\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\beta}_\theta, \sigma_\theta^2, \boldsymbol{B}, \boldsymbol{\beta}_b, \sigma_b^2, \boldsymbol{G}, \boldsymbol{\beta}_g, \sigma_g^2, \boldsymbol{\delta}, \sigma_\delta^2, \boldsymbol{Z}, \pi | \boldsymbol{Y}) &\propto \prod_{i=1}^{N} \prod_{k=1}^{M} \mathrm{Cat}(\boldsymbol{Y} | \boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{G}, \boldsymbol{\delta}, \boldsymbol{Z}) \\
&\times \prod_{i=1}^{N} N(\theta_i | \boldsymbol{\beta}_\theta, \sigma_\theta^2) \prod_{i=1}^{N} N(\mathrm{logit}(g_i) | \boldsymbol{\beta}_g, \sigma_g^2) \prod_{i=1}^{N} N(\mathrm{logit}(b_i) | \boldsymbol{\beta}_b, \sigma_b^2) \\
&\times \prod_{k=1}^{M} N(\delta_k | 0, \sigma_\delta^2) \prod_{k=1}^{M} \mathrm{Bern}(Z_k | \pi) \\
&\times \prod_{f=1}^{F} N_{J+1}(\boldsymbol{\beta}_f | 0, 10\mathbf{I}_{J+1}) \prod_{h=1}^{H} U(\sigma_h | 0.01, 4) U(\pi | 0, 1),
\end{aligned} \tag{17}$$

where the first expression after the proportionality sign is the likelihood, described in detail in equation (6). It is followed by the products of the population densities of the person-specific parameters as specified in equations (7), (11), and (12). The next line describes the population densities of the item-specific parameters given their population mean and variance parameters, as in equations (8) and (13). Finally, the last line multiplies all the above by the prior densities, as chosen in equations (14), (15), and (16).

As mentioned above, we rely on the computer software JAGS to draw samples of the conditional posterior distribution. MATLAB was used for calling JAGS and interpreting its output. Program scripts are available as online supplements at the Web site of *Sociological Methodology*. In the next section, the merits of the HECM are demonstrated by fitting it to 2010 GSS and 1994 IAA-FAQ data.

6. APPLYING THE HECM TO SURVEY DATA

In the next subsections, the HECM is applied to two knowledge areas: scientific knowledge and knowledge on aging. In both cases, we use standard sets of questions that have been widely studied. Obviously these are not the only topical areas that are suitable for CCT, and we return to possible extensions in the Discussion.

6.1. *The 2010 GSS Science Questions*

Within a survey context, standard batteries of science knowledge questions have been developed and refined for many decades (Miller 1998; Bann and Schwerin 2004). Although there is some variation in the exact questions included in particular surveys, most of the questions are true-or-false questions of scientific concepts that offer DK options. The 2010 GSS (in collaboration with the National Science Foundation) featured in this application section uses a set of 12 questions, 11 of which are true-or-false format with DK options provided. Appendix A lists the GSS questions. In addition, there are several other variables that can be used as predictors, five of which were chosen for covariate analysis for the current application: education (in terms of highest grade), gender (1 = female, 2 = male), age, religiousness (1 = very religious, 4 = not religious), and whether respondents have had religious experiences that changed their lives (1 = yes, 2 = no).

Although the ECM can be applied to knowledge areas without a predefined answer key, science knowledge provides an interesting application area for CCT because the scientifically correct answers are known a priori (Allum et al. 2008; Durant, Evans, and Thomas 1989; Miller 1998, 2004; National Science Board 2012; Pardo and Calvo 2004). With CCT analysis, a cultural consensus answer key based on patterns of respondents' answers is derived and compared with the scientifically correct answers. Generally speaking, no particular objective truth value is attached to the derived consensus answers, save that it is a property of the group of respondents. It is important to distinguish between experts' knowledge (e.g., what scientists agree about various phenomena) and the possibility that some respondents share nonscientific consensus, keeping in mind that from the perspective of cultural consensus, "knowledge" refers to the shared understandings among members of some population with respect to a particular domain. Previous research has indicated that adults in the United States and elsewhere exhibit relatively modest amounts of knowledge when asked about scientific concepts. On possible bases for low scientific literacy, see Miller (2010a, 2010b, 2011). For example, the 2010 National Science Foundation study found that adults in the United States correctly answered 63 percent of 12 science knowledge questions (National Science Board 2012). Although these simple aggregation-based summary measures are designed primarily to measure scientific ignorance, CCT is able to detect counter-knowledge: a consensus-based rejection of certain scientific ideas. Detecting counter-knowledge is important because it can lead to recommendations in policy making, among other things.

Science knowledge can differ among members of different subcultures or social groups. Prior research on science literacy has found higher levels of knowledge among better educated people and those who have taken more college science courses (Miller 2004; National Science Board 2012) and among younger people (Miller 2004; National Science Board 2012). Often there is a detectable gender gap in knowledge; namely, women are found to be less knowledgeable (Hayes and Tariq 2000; Miller 2004), though much of the gap can be explained by gender differences in the willingness to guess (Mondak and Canache 2004). There is also evidence that people who hold a literal interpretation of the Bible are less knowledgeable (Miller, Scott, and Okamoto 2006; Zigerell 2012). These results informed our selection of covariates for the HECM.

The HECM can contribute to the investigation of science knowledge because many of the issues raised in the study of science knowledge are directly related to aspects of CCT and to the parameters of the hierarchically extended model: item difficulty (Miller 1998; Bann and Schwerin 2004; National Science Board, 2012), acquiescence bias (Pardo and Calvo 2004), the effects of guessing (Miller 1998; Mondak and Canache 2004), the use of DK responses (Bauer 1996; Mondak and Canache 2004; Sturgis et al. 2008), factors that are associated with variation in levels of knowledge, and the possibility that there is cultural variability in understandings of science (Miller 2010a, 2010b; Pardo and Calvo 2004).

The 2010 GSS included $M = 12$ items on scientific knowledge. The data have 927 respondents who answered these items. Eleven respondents were removed because they had missing explanatory variables; hence, the final sample size is $N = 916$. Table 1 shows summary

Table 1. Descriptive Statistics and Results on the Questionnaire Items with the HECM, Based on the 2010 GSS Data

Item Keywords	Raw Data: True/False/DK	Scientific Answers	Scored Answers	Majority Answers	HECM Answer Key	HECM ID
1. Earth's center is hot	759/ 55/102	True	True	True	True	–2.11
2. Radioactivity man-made	168/612/136	False	False	False	False	–0.56
3. Father's gene	564/223/129	True	True	True	True	0.25
4. Laser—sound wave	175/434/307	False	True	False	False	1.25
5. Electrons < atoms	474/181/261	True	True	True	True	0.73
6. Antibiotics kill viruses	423/441/ 52	False	True	False	False	0.91
7. Universe began explosion	340/307/269	True	False	True	True	2.15
8. Continents move	730/ 85/101	True	True	True	True	–1.68
9. Humans from animals	447/351/118	True	False	True	True	1.13
10. Earth around sun	687/163/ 66	True	True	True	True	–1.38
11. Tomatoes no genes	180/439/297	False	True	False	False	1.22
12. Cloning identical	739/ 79/ 98	True	True	True	True	–1.82

Note: A full description of items can be found in Appendix A. The ''Scientific Answers'' column displays the correct answers established by scientific experts. The ''Scored Answers'' column shows the majority answers when DK responses are counted toward the incorrect response alternative. The ''Majority Answers'' column shows whether the majority of the answer is ''true'' or ''false,'' while ignoring the DK responses (treating them as missing data). The HECM answer key displays the posterior median estimates from the HECM, with 1's labeled as ''true'' and 0's as ''false.'' The last column shows the posterior mean estimate of the IDs. DK = don't know; HECM = hierarchical extended Condorcet model; ID = item difficulty.

statistics on the items. The first column denotes the items by keywords. The next column shows how many people answered ''true,'' ''false,'' or ''don't know'' to these questions. As can be seen, for some items there are a large number of DK responses. It is important to note that the HECM works with the complete $N \times M$ data of trichotomous responses, not only with the marginals, as displayed in the second column of Table 1.

Column 3 displays the correct answers to these questions, as established by scientific experts. There are two ways to construct simple tallies of correct and incorrect responses. On the basis of the ''scientific answers,'' the DK responses are often interpreted as incorrect answers in assessment, and they are counted toward the incorrect response alternative. Column 4 in Table 1 displays the ''scored answers,'' which are the majority responses with DK responses scored toward the incorrect alternative. Alternatively, we can calculate ''majority answers'' by interpreting DK responses as missing data and checking which answer (''true'' or ''false'') was provided by more respondents, as shown in column 5. This approach is rather questionable because more than 17 percent of the responses were DK responses. As can be seen, there are five items for which coding the DK responses as incorrect responses would result in ''scored answers'' that are different from the ''majority answer'' as well as from the ''scientific answer.'' By resorting to simple summary statistics, we neglect to take into account interindividual differences in cultural knowledge. This is because when these answer keys are established, all respondents count equally toward the final decision of whether an item is regarded ''true'' or ''false'' in the population, and interindividual differences in response style (e.g., in the willingness to guess) are also disregarded. In contrast, the HECM explicitly models the decision process by individual-specific parameters when deriving the consensus answer key.

6.2. *Results on the 2010 GSS Data with the HECM*

The HECM was fit to the data described above.[2] The estimated consensus answer key is shown in column 6 in Table 1. This answer key is based on the posterior median estimates calculated from the Bernoulli conditional posterior distribution for each item. The median estimates are either 1 or 0, and these values are labeled as ''true'' or ''false'' in column 6. The posterior median was equal to the posterior mean for every answer key estimate up to two decimal points. In the Bernoulli distribution, the mean is equal to the median only if all values are the same and close to it only if there is low variability in the sample chain, indicating 0 posterior standard deviation. Therefore this result indicates a high degree of certainty in the estimates.

As can be seen in columns 3 and 6 in Table 1, the consensus answers are identical to the expert answers. We would like to emphasize that in the HECM, the DK answers are treated as legitimate responses rather than disregarded as missing data or treated as incorrect responses. In our analysis, the DK responses were taken into account as part of the cognitive decision process. The positive message of the results is that on the basis of the presented cognitive model, the population shows similar consensus answers on the scientific items as the expert answers. We emphasize the fact that the HECM answer key was estimated solely from the data while assuming noninformative prior distributions on the model parameters, and the established expert answer key was not entered in any way into the estimation routine.

The population can be described in terms of the estimates of the posterior mean and variance of the HECM parameters, as shown in Table 2. By checking the population variance ($\sigma_\theta^2 = 1.69$) of the ability parameter, we can conclude that there is substantial interindividual variation in ability. The average ability is –0.61. As described earlier in equation (5), the ability together with the item difficulty determine how likely it is that a person knows the correct answer to an item (D_{ik}; see Figure 3 for the function of this parameter in the decision process). As the mean of the item difficulty is fixed to 0, if we input the average ability into equation (5), it can be seen that on average the probability that a respondent knows the correct answer is 0.35. The average ability can also be interpreted in connection with each item's difficulty estimate (δ_k, in the last column of Table 1) to investigate which items are most likely to be known by a respondent with average ability. Generally speaking, the population level results suggest that items with difficulty less than –0.61 result in a higher than chance probability of being known ($D_{ik} > 0.5$) by an average respondent. As can be seen from Table 1, there are four items like that.

The next four rows of Table 2 describe the respondents in terms of cognitive decision process variables that are independent of their abilities. From the population mean of willingness to guess, we see that respondents on average are rather willing to guess; however, there is

Table 2. Population Level Results Based on the HECM (2010 GSS Data)

Model Parameter	Parameter Description	Posterior Mean	CI Percentiles 2.5%	97.5%
μ_θ	Population mean—ability	–0.61	–1.57	0.36
σ^2_θ	Population variance—ability	1.69	1.31	2.14
μ_b	Population mean—willingness to guess (logit scale)	1.12	1.01	1.24
σ^2_b	Population variance—willingness to guess (logit scale)	1.51	1.22	1.85
μ_g	Population mean—guessing bias (logit scale)	–0.05	–0.17	0.08
σ^2_g	Population variance—guessing bias (logit scale)	0.55	0.37	0.77
σ^2_δ	Population variance—item difficulty	2.89	1.13	6.96
π	Answer key "true" probability	0.64	0.39	0.86

Note: CI = credible interval; GSS = General Social Survey; HECM = hierarchical extended Condorcet model.

some interindividual variation here as well, represented by the population variance parameter for willingness to guess. The guessing bias appears to be fairly neutral on average, meaning that respondents in general do not prefer to answer "true" rather than "false" (or vice versa) when they do not know the correct answer.

The last two rows of Table 2 concern the 12 science knowledge items. To begin with, these items show a rather large variation in their item difficulty levels. From Table 1, it turns out that the most difficult question is "The universe began with a huge explosion," and the easiest is "The center of the Earth is very hot." The findings on the item difficulty are in line with previous results of fitting IRT models to earlier years of the GSS science items (Miller 1998, 2004; Bann and Schwerin 2004). Finally, the last line of Table 2 shows that it is more likely that a latent answer key item is "True" rather than "false" (π = 0.64) for this set of questions.

With respect to the explanatory variables, we included the respondents' gender, age, years of education, religiousness (self-assessed), and

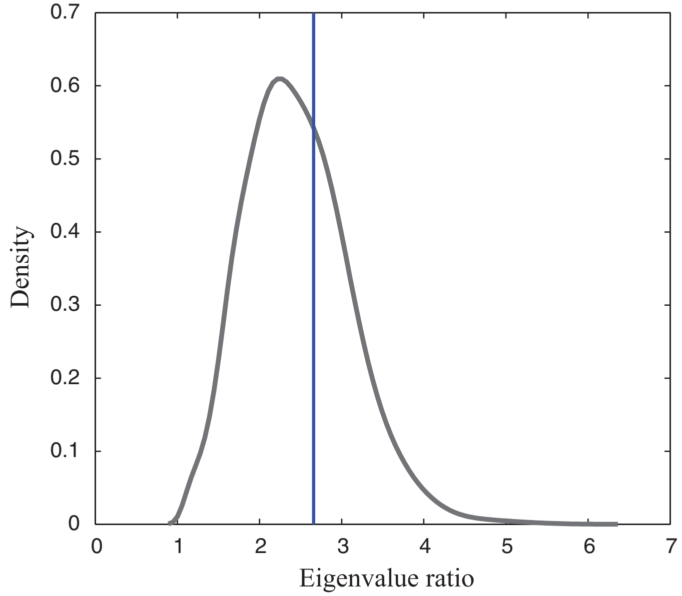

Figure 3. Eigenvalue ratio of the posterior predictive distribution and the 2010 General Social Survey data.

whether they have ever had a religious life-changing experience. This covariate set was chosen because these variables have been widely discussed in connection with scientific knowledge (see previous discussion).[3] Religiousness has been related to science knowledge through specific items, namely, questions about humans developing from animals and the universe starting with an explosion (National Science Board 2006, 2008). However, the religiousness predictor is included in our model to measure more general tendencies.

Table 3 displays the results. In the Bayesian sense, a predictor has a credibly nonzero regression coefficient (the word ‘‘significant’’ is not used in the Bayesian context) if its 95 percent credible interval does not contain 0. Results related to ability are generally consistent with prior research (e.g., see Hayes and Tariq 2000; Miller 2004; Miller et al. 2006; National Science Board 2012): educated people had higher ability, religious people had lower ability, and the results showed a gender gap in science knowledge. However, a previously observed age effect

Table 3. Results on the Covariates Based on the HECM (2010 GSS Data)

			CI Percentiles	
Model Parameter	Covariate	Posterior Mean	2.5%	97.5%
Ability	Education	1.16	1.01	1.32
	Gender	–0.29	–0.41	–0.16
	Age	–0.05	–0.18	0.07
	Religiousness (‘‘very’’ to ‘‘not’’)	0.40	0.27	0.53
	Religious experience changed life	0.04	–0.10	0.17
Willingness to guess	Education	0.02	–0.09	0.14
	Gender	–0.24	–0.36	–0.13
	Age	–0.22	–0.33	–0.11
	Religiousness (‘‘very’’ to ‘‘not’’)	–0.11	–0.24	0.01
	Religious experience changed life	–0.18	–0.30	–0.07
Guessing bias	Education	–0.27	–0.37	–0.16
	Gender	0.04	–0.05	0.14
	Age	–0.02	–0.11	0.07
	Religiousness (‘‘very’’ to ‘‘not’’)	0.13	0.03	0.24
	Religious experience changed life	0.08	–0.02	0.17

Note: CI = credible interval; GSS = General Social Survey; HECM = hierarchical extended Condorcet model.

(namely, that younger people are generally more knowledgeable than older people) was not found by the current analysis. With respect to the other variables of the decision process, men and younger people tended to guess more often when they were unsure. Also, the results showed that respondents who had religious life-changing experiences were more willing to guess. When it comes to guessing, educated and religious respondents guessed ‘‘false’’ more often.

The conclusion of this analysis is that CCT-based consensus knowledge on the 12 science knowledge items corresponds to the expert knowledge. Covariate analysis found a gender gap in science knowledge even when response style tendencies were accounted for. That suggests that although women respondents were indeed less willing to guess, they also tended to exhibit lower levels of knowledge in the science domain, even after a gender difference in guessing tendency was taken into account. When it comes to religious respondents, it appears that they also had lower level of knowledge in terms of GSS questions, and those with life-changing religious experiences were more likely to guess.

6.3. *Palmore's FAQ*

Our second empirical application uses Palmore's FAQ (Palmore 1977, 1998), a test that was ''designed to cover the basic physical, mental, and social facts and the most common misperceptions about aging'' (Palmore 1977). The quiz contains 25 factual statements about older people, with ''true,'' ''false,'' and ''don't know'' response options (see Appendix B). The version analyzed here was administered as part of the IAA survey (American Association of Retired Persons 1994). This version incorporates suggested revisions to the original FAQ; specifically, it includes a DK option (American Association of Retired Persons 1994) and clarifies ambiguous terms (Miller and Dodder 1980). Studies suggest that the DK option reduces guessing and provides a more accurate measure of knowledge (Miller and Dodder 1980; Courtenay and Weidemann 1985; Seufert and Carrozza 2002). Although all of the questions in the FAQ are empirically substantiated (Palmore 1977), two subsets of questions can be defined, which, if answered incorrectly, indicate either positive or negative biases toward older people (Palmore 1977).[4]

The FAQ has been widely used in the decades since its introduction (Palmore 1980, 1982, 1998, 2005), often in instructional settings (e.g., see Harris and Dollinger 2001; Stuart-Hamilton and Mahoney 2003). It has also been subject to a number of evaluations concerning its reliability and validity (Klemmack 1978; Palmore 1980; Norris, Tindale, and Matthews 1987), its factor structure (Klemmack 1978; Norris et al. 1987; Palmore 1978), and how well the quiz serves as either an indicator of knowledge or of attitudes toward older people (Courtenay and Weidemann 1985; Holtzman and Beck 1979; Klemmack 1978; Stuart-Hamilton and Mahoney 2003; Palmore 1978, 1982, 2005).

With respect to cultural consensus on knowledge on aging, we view this as a potentially fruitful example because the FAQ highlights how survey questions on a knowledge domain can combine factual knowledge with socially constructed stereotypes and attitudes—in this case, the generally negative stereotypes of older people (Harris and Dollinger 2001; Palmore 2005).

The IAA (American Association of Retired Persons 1994) surveyed 1,200 respondents to identify some of the social and economic concerns related to aging. In addition to Palmore's FAQ, the survey included many other variables, four of which are used in the following analysis as covariate information: education (1 = grade school or less, 7 = post-graduate), gender (1 = female, 2 = male), age, and frequency of contact with older people (1 = daily, 7 = never). The current analysis used all $M = 25$ items from Palmore's questionnaire. The sample size is $N =$ 1,167 (33 respondents were removed because they had missing explanatory variables). Table 4 shows summary statistics on the items. Column 1 denotes the items by keywords. Column 2 shows the responses summed over response categories. Column 3 displays the correct answers to these questions, as established by gerontologists (Palmore 1977). Column 4 displays these ''scored answers'' (calculated the same way as in the 2010 GSS application) in terms of ''true'' or ''false'' for every item. Column 5 shows the ''majority answers'' in terms of ''true'' or ''false'' while interpreting DK responses as missing data. As can be seen, there are three items for which coding the DK respnoses as incorrect responses would result in ''scored answers'' that are different from the ''majority answers.''

6.4. *Analyzing Data from Palmore's FAQ*

The HECM was fit to the data described above. Results on the items are shown in Table 4. Posterior median estimates (labeled ''true'' and ''false'') are displayed in column 6. As in the previous application, the answer key items estimates showed posterior standard deviations of essentially 0, indicating a high degree of confidence in the answer key estimates obtained from the model. These results are in correspondence with passing the one-culture test (see later), which suggest that there is consensus among the respondents on aging. Finally, column 7 shows the item difficulty estimates on the basis of posterior means.

In this analysis, there are nine items for which the consensus answers do not coincide with the expert answer key, namely, 7, 8, 12, 16 to 18, 20, 24, and 25. Of these nine questions, eight indicate negative age bias and one suggests positive age bias (see again note 4). In fact, seven of these nine question fall in this category: 7, 8, 16 to 18, 24, and 25, while

Table 4. Descriptive Statistics and HECM Answer Key Estimates on Palmore's Quiz

Item Keywords	Raw Data: True/False/DK	Scientific Answers	Scored Answers	Majority Answers	HECM Answer Key	HECM ID
1. Senile, defective memory	147/976/ 44	False	False	False	False	–1.75
2. All five senses decline	875/241/ 51	True	True	True	True	0.14
3. No capacity for sex	164/869/134	False	False	False	False	–1.05
4. Lung capacity declines	782/218/167	True	True	True	True	1.04
5. Miserable	271/816/ 80	False	False	False	False	–0.74
6. Physical strength declines	1,094/59 /14	True	True	True	True	–2.24
7. One-tenth institutionalized	817/173/177	False	True	True	True	0.73
8. Fewer driving accidents	481/501/185	True	False	False	False	1.25
9. Cannot work as effectively	330/766/ 71	False	False	False	False	–0.49
10. Carry out normal activities	917/188/ 62	True	True	True	True	–0.47
11. Unable to adapt to change	520/591/ 56	False	False	False	False	0.46
12. Take longer to learn new	575/516/ 76	True	False	True	False	1.06
13. Impossible to learn new	76/1,080/ 11	False	False	False	False	–2.75
14. React more slowly	940/165/ 62	True	True	True	True	–0.43
15. They are alike	339/799/ 29	False	False	False	False	–0.67
16. Seldom bored	359/723/ 85	True	False	False	False	–0.25
17. Lonely	775/290/102	False	True	True	True	1.26
18. Fewer work accidents	566/408/193	True	False	True	False	2.36
19. Over 20% in the United States	793/130/244	False	True	True	True	0.61
20. Low priority for doctors	413/615/139	True	False	False	False	0.41
21. Incomes below poverty	535/443/189	False	False	True	False	1.55
22. Working or would like to	1,062/75/ 30	True	True	True	True	–1.79
23. Become more religious	756/289/122	False	True	True	True	1.39
24. Seldom angry	368/688/111	True	False	False	False	–0.04
25. Will have worse status in 2020	836/240/ 91	False	True	True	True	0.48

Note: A full description of items can be found in Appendix B. The ''Scientific Answers'' column displays the correct answers established by gerontologists. The ''Scored Answers'' column shows the aggregated answer in the data when DK answers are counted toward the incorrect response alternative. The ''Majority Answers'' column shows whether the majority answers are true or false while ignoring the DK responses. The HECM answer key shows the posterior median estimates from the HECM (with ''true'' coded 1 and ''false'' as 0).The last column shows the posterior mean estimate of the IDs. DK = don't know; HECM = hierarchical extended Condorcet model; ID = item difficulty.

question 12 actually suggests some positive bias toward the elderly. Although this lack of correspondence with the expert answer key appears to indicate a tendency for negative age bias, the difference in the proportions of ''incorrect'' answers to the negative bias and positive bias question sets is not significant ($p = .338$, Fisher's exact test).

In addition, there are three items (12, 18, and 21) for which the HECM answer key deviates from the majority answer key. The true/false/DK ratio in the raw data indicates moderate split among the respondents for these three items. The combination of the split and relatively high number of DK responses for these items leads to high item-difficulty levels of 1.06, 2.36, and 1.55 respectively. High item difficulties imply that many people do not know the correct answer, allowing guessing to play a considerable part in their final responses. Because in our framework, guessing and willingness to guess are accounted for, the answer key estimates obtained from the HECM are most likely dominated by the answers of the most knowledgeable individuals, who for these items go against the majority's opinion.

Table 5 displays the population-level results. The overall consensus knowledge level was higher in this application than in the previous one: The population mean for knowing the consensus answer is –0.38, which translates into about 0.41 in the unit probability scale. The interindividual variation in this ability is not very high, $\sigma^2_\theta = 0.44$. The respondents in general were quite willing to guess, with probability of guessing of about 0.92, although there was large interindividual variation in this ($\sigma^2_b = 2.54$). Also, the population mean of the guessing bias suggests a general tendency to guess ''true,'' with some interindividual variation. Finally, the last row of the table indicates that on average the items were almost as likely to be ''true'' as ''false.''

Table 6 summarizes results on the covariate modeling. As before, a predictor is considered to have a credibly nonzero regression coefficient if its 95 percent credible interval does not contain 0. Although ability was related only to education, willingness to guess was influenced by education, gender, and age as well: More educated, women, and older respondents were less willing to guess. Finally, it was shown that contact with elderly people and age made respondents more likely to guess ''true'' on the items. In contrast, educated respondents were connected to guessing ''false.''

Table 5. Population Level Results Based on the HECM (1994 IAA-FAQ Data)

Model Parameter	Parameter Description	Posterior Mean	CI Percentiles 2.5%	CI Percentiles 97.5%
μ_θ	Population mean—ability	–0.38	–0.91	0.15
σ^2_θ	Population variance—ability	0.44	0.34	0.55
μ_b	Population mean—willingness to guess (logit scale)	2.41	2.29	2.54
σ^2_b	Population variance—willingness to guess (logit scale)	2.54	2.18	2.95
μ_g	Population mean—guessing bias (logit scale)	0.76	0.66	0.86
σ^2_g	Population variance—guessing bias (logit scale)	0.43	0.34	0.53
σ^2_δ	Population variance—item difficulty	1.85	1.01	3.35
π	Answer key—"true" probability	0.45	0.27	0.63

Note: CI = credible interval; FAQ = Facts on Aging Quiz; HECM = hierarchical extended Condorcet model; IAA = Images of Aging in America.

Table 6. Results on the Covariates Based on the HECM (1994 IAA-FAQ Data)

Model Parameter	Covariate	Posterior Mean	CI Percentiles 2.5%	CI Percentiles 97.5%
Ability	Education	0.31	0.25	0.37
	Gender	–0.01	–0.07	0.05
	Age	–0.04	–0.10	0.03
	Contact with elderly (often to never)	0.01	–0.06	0.07
Willingness to guess	Education	–0.21	–0.33	–0.10
	Gender	–0.19	–0.30	–0.08
	Age	–0.13	–0.25	–0.02
	Contact with elderly (often to never)	–0.01	–0.13	0.10
Guessing bias	Education	–0.20	–0.27	–0.14
	Gender	–0.02	–0.08	0.04
	Age	0.08	0.02	0.14
	Contact with elderly (often to never)	0.09	0.03	0.15

Note: CI = credible interval; FAQ = Facts on Aging Quiz; HECM = hierarchical extended Condorcet model; IAA = Images of Aging in America.

7. MODEL CHECKING

A crucial assumption of the HECM is that there is one underlying culture behind the data. Although these checks should always be carried out before interpreting the results, we present them at this point to be able to look at the two data sets at once.

In Bayesian statistics, absolute model fit is generally investigated through posterior predictive model checks (PPCs). They are carried out by first selecting statistics (PPC test statistics) that reflect important features of the real data. Then, on the basis of the posterior distribution of the model parameters and the model itself, several hundred (or thousand) data sets are generated, and the PPC test statistics are calculated for each of these replicated data sets. The same statistics are then calculated for the real data. If the real data statistics do not appear to be consistent with the distribution of statistics generated from the replicated data, it is unlikely that the proposed model provides a good description of the real data.

Batchelder and Anders (2012) developed a PPC to test the one-culture assumption. Their method is based on calculating the respondent-by-respondent correlations on the basis of the respondents' answers (except the DK answers in the case of the HECM), calculating eigenvalues from the correlation matrix, and computing the ratio of the first and second eigenvalues. Such a measure is analogous to the indicator of a one-factor solution in factor analysis—namely, it tests whether the first factor accounts for most of the variation among the correlated variables and the other factors are simply fitting noise in the data. It is thus expected that the first eigenvalue is a multiple of the second one, although the actual multiplying constant is a property of the data scale.

We implement a PPC on the basis of the eigenvalue measure not by simply looking at the absolute value of eigenvalue ratio in the data but by comparing it with a posterior predictive distribution based on generated data sets. When data sets are generated on the basis of the parameter estimates and the HECM, the one-culture assumption is automatically met, because it is a property of the model from which the data are generated. Therefore, the ratio of the first and second eigenvalues that are based on these generated data sets represents a possible range of eigenvalue ratios with a single underlying culture. A 95 percent posterior predictive credible interval on the basis of the generated

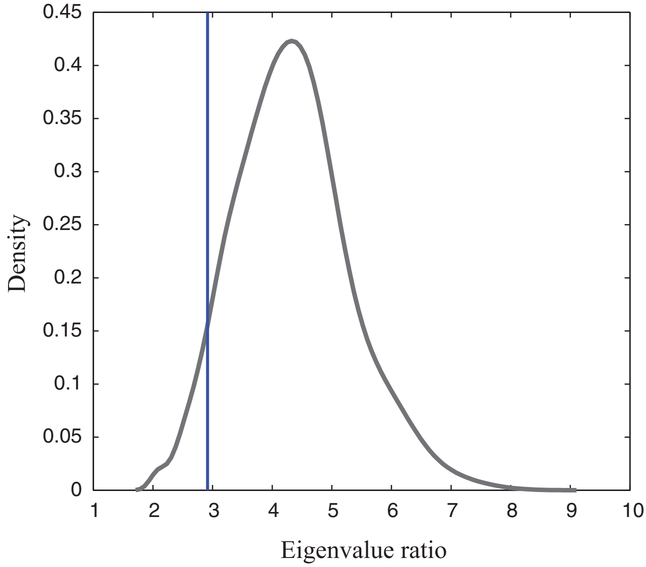

Figure 4. Eigenvalue ratio of the posterior predictive distribution and 1994 Images of Aging in America and Facts on Aging Quiz data.

eigenvalue ratios should contain the eigenvalue ratio calculated from the real data.

However, we note here that if the data sets are based on all distributional assumptions of the HECM, this posterior predictive check is therefore also sensitive to these properties. Simulation studies presented in Batchelder and Anders (2012) and Anders and Batchelder (2012) show the efficiency of this PPC to detect deviations from the one-culture assumption.

Figure 3 displays graphically the posterior predictive check results on the 2010 GSS data. The first and second eigenvalue ratio was calculated in 300 simulated data sets, and a smoothed histogram of these values is shown in Figure 3. The straight line represents the eigenvalue ratio in the 2010 GSS data. As can be seen, the line falls almost in the middle of the posterior predictive distribution, supporting the fact that the single-culture assumption holds for the 2010 GSS data.

Figure 4 shows the same test carried out for the 1994 IAA-FAQ data. As can be seen, the eigenvalue ratio for the data lies on the far left side of the posterior predictive distribution; however, the numerical results confirm that this data eigenvalue ratio is within the 95 percent posterior predictive credible interval.

8. ROBUSTNESS OF THE HECM ANSWER KEY ESTIMATES

The current application of CCT for social survey data with large samples is a somewhat untraditional CCT application because it has large numbers of respondents but small numbers of items. Typically in CCT applications, there are at least 30 items tapping into the same knowledge or belief system, and there are only a couple of dozen respondents, sometimes even fewer. It has been shown that even with only a handful of reasonably knowledgeable respondents (which can be a number as low as 6), the answer key can be recovered accurately (e.g., see Batchelder and Anders 2012).

To test the robustness of the findings for the HECM, we can sample random subsets from the data set, reestimate the HECM, and compare the answer key estimates among the subsamples. We choose the 2010 GSS data set to demonstrate this test.

Three subsample sizes were chosen (N = 25, 50, and 100), and 10 random subsamples were selected for each. The results are displayed in Table 7. Columns 2, 3, and 4 show how many times the subsamples' answer key estimates matched the ones estimated on the full sample. Already with the smallest subsample size (N = 25), the findings appear to be robust for many items, and with the largest sample size (N = 100), all answer key estimates are the same among the subsamples and the full sample.

9. DISCUSSION

In this article, we have described and illustrated the HECM for CCT, and we have demonstrated its value for analyzing knowledge questions in social surveys. The model generalizes the GCM by allowing uncertainty in responses (DK answers), extending it hierarchically, and introducing covariates. These extensions provide insights into theoretically important aspects of cultural knowledge, including systematic variation in levels of knowledge, tendencies to guess, and acquiescence bias.

Table 7. Robustness Test for 2010 GSS

Item Keywords	$N = 25$	$N = 50$	$N = 100$
Earth's center is hot	10	10	10
Radioactivity man-made	8	10	10
Father's gene	9	10	10
Laser—sound wave	8	9	10
Electrons < atoms	10	10	10
Antibiotics kill viruses	7	8	10
Universe began explosion	8	9	10
Continents move	10	10	10
Humans from animals	9	9	10
Earth around sun	10	10	10
Tomatoes no genes	7	9	10
Cloning identical	10	10	10

Note: Three times, 10 random subsamples of the original data set were taken, with sample sizes of $N = 25$, $N = 50$, and $N = 100$. Columns 2, 3, and 4 display how many times the subsamples' answer key estimates matched the one based on the full sample.

Taking into account cognitive processes underlying decision making as well as individual differences therein can contribute to our understanding of sociological phenomena. Researchers such as Mondak and Canache (2004) have already called our attention to the importance of these factors. The authors argued that individual and national differences in tendency to guess can complicate measures of knowledge, because respondents who are more likely to guess will get some answers correct simply by chance. The HECM allows direct examination of such hypotheses.

One of the advantages of CCT modeling is that its underlying cognitive model separates "knowledge" from "performance," using a latent variable for each. Performance can be defined by how many items a respondent answers correctly, in which case guessing can lead to correct responses. By looking again at Figure 4, we can see that the underlying cognitive model in the HECM works not with performance but with the underlying latent probability of actually knowing the answer, denoted by D_{ik}. By taking into account willingness to guess and guessing bias, and allowing these parameters for each individual to take different values, we account for those "correct" responses that are not based on the respondent's knowledge.

As we saw in the results for the FAQ, HECM answer key estimates can differ from the scientific answers and from the majority answers as well (see again Table 4). An HECM answer key can differ from a simple aggregation-based majority answer because it accounts for guessing processes and interindividual differences therein, and it considers the true underlying probability of a respondent's knowing the consensus answer for a particular question. Therefore, the HECM answer key conveys an interpretation of consensus opinion in society that is based on principles of cognitive and personality psychology. In some domains, there are empirical evidence and accepted scientific methods that yield the scientifically correct answers. These scientific answers can vary from the HECM key if shared cultural knowledge differs from that established by the scientific community. Discrepancies could arise if cultural knowledge is out of date, for example, if a substantial portion of a population learned about mountain orogeny before plate tectonics was fully understood. It can also arise if worldviews concerning particular questions are embedded in other belief domains, for example, fundamentalist religious beliefs about evolution or the age of the universe. In either case, discrepancies can point to lines of further investigation (Why are there discrepancies?) or suggest educational interventions (What needs to be taught?). Finally, we note here that there are cases for which no scientifically ''correct'' answer exists (e.g., morality judgments, meaning of culturally specific expressions).

Although CCT has, to date, seen little use in sociology, we believe that it has considerable potential for systematically examining a number of issues in the field. Knowledge domains are one natural area for further exploration, and the model could easily be applied to other topics for which survey data are readily available, such as political facts (Carpini and Keeter 1993; Mondak 2001; Mondak and Anderson 2004), as well as more everyday knowledge, such as diet and health (York-Crowe et al. 2006). Application of CCT could also be expanded to include more attitudinal or value-oriented questions, for example, on human rights, the environment, morality, or identity. Moreover, the availability of comparable cross-national surveys (e.g., World Values Survey Association 2009) makes this an especially promising area of investigation for CCT.

Nevertheless, several challenges remain in the application of CCT to large-scale survey data. First, as we discussed in the previous section, CCT was originally developed for situations in which a handful of respondents answer a relatively large number of questions. Survey data with many respondents have high computational costs, especially when conducting posterior predictive checks, for example. Second, because large representative surveys often have considerable respondent heterogeneity, the one-culture assumption of CCT might not always be met in practice. Future research should consider principled ways to model heterogeneity in cultural knowledge, for example, implementing multicultural models, as in Anders and Batchelder (2012) for social survey research as well. Finally, the model described in this article is applicable to questions with dichotomous response formats (true or false) plus a DK option. However, surveys commonly use ordered categories for responses (''strongly agree,'' ''agree,'' ''neither agree nor disagree,'' ''disagree,'' ''strongly disagree''). Extending CCT models to such formats will certainly increase their applicability.

APPENDIX A

The 2010 General Social Survey Science Module

1. The center of the Earth is very hot.
2. All radioactivity is man-made.
3. It is the father's gene that decides whether the baby is a boy or a girl.
4. Lasers work by focusing sound waves.
5. Electrons are smaller than atoms.
6. Antibiotics kill viruses as well as bacteria.
7. The universe began with a huge explosion.
8. The continents on which we live have been moving their locations for millions of years and will continue to move in the future.
9. Human beings, as we know them today, developed from earlier species of animals.
10. Does the Earth go around the Sun, or does the Sun go around the Earth?
11. Ordinary tomatoes do not contain genes, while genetically modified tomatoes do.
12. The cloning of living things produces genetically identical copies.

APPENDIX B

Palmore's Aging Quiz

1. The majority of older people are senile, have a defective memory, or are disoriented.
2. All five senses (sight, smell, hearing, taste, touch) tend to decline in old age.
3. The majority of older people have no capacity for sexual relations.
4. Lung capacity tends to decline in old age.
5. The majority of older people say they are miserable most of the time.
6. Physical strength tends to decline in old age.
7. At least one-tenth of the older people are living in institutions such as nursing homes, mental hospitals, homes for the aged.
8. Drivers 65 or older have fewer accidents per driver than those under age 65.
9. Older workers usually cannot work as effectively as younger workers.
10. Over three-fourths of older people say they are healthy enough to carry out their normal activities.
11. The majority of older people are unable to adapt to change.
12. Older people tend to take longer to learn something new.
13. It is almost impossible for older people to learn something new.
14. Older people tend to react more slowly than younger people.
15. In general, older people tend to be pretty much alike.
16. The majority of older people say they are seldom bored.
17. The majority of older people say they are lonely.
18. Older workers have fewer accidents than younger workers.
19. Over 20% of the U.S. population are now 65 and older.
20. The majority of medical practitioners such as medical doctors and nurses give low priority to older people.
21. The majority of older people have incomes below the poverty level—that is less than $6,500 per year for an individual or less than $8,400 per year for a couple.
22. The majority of older people are working or say they would like to have some kind of work to do, including work around the house and volunteer work.
23. Older people tend to become more religious as they age.
24. The majority of older people say they are seldom irritated or angry.

25. The health and economic status of older people in the year 2020 will be probably the same or worse than it is now.

Acknowledgments

We would like to thank the five anonymous reviewers for their constructive comments.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based in part on work supported by the U.S. Army Research Office under grant FA9550-09-1-0510.

Notes

1. Alternatively, the population mean of the ability parameter could also be fixed to 0.
2. In the analyses of the 2010 GSS data set as well as the 1994 IAA-FAQ data set, three chains were run with dispersed starting values, and 4,000 iterations were retained from each chain after an initial burn-in period (2,000 samples) and thinning (by factor 10). Therefore, the final posterior sample size was $3 \times 4{,}000 = 12{,}000$ iterations. All chains converged. The primary test for convergence was based on the $\hat{R}$ value (i.e., all $\hat{R}$ values < 1.1; Gelman et al. 2004), as implemented in the MATLAB environment by the authors, and visual assessment of trace and cumulative plots. Also, other tests of convergence, such as Geweke and Heidelberger and Welch diagnostics, were carried out satisfactorily on the basis of the R package superdiag by Tsai and Gill (2012).
3. Although Mondak and Canache (2004), for example, demonstrated that there is a gender difference in the tendency to guess, they only approximated a guessing rate as (number incorrect)/(number incorrect + number DK).
4. Questions 1, 3, 5, 7 to 11, 13, 16 to 18, 21, 22, 24, and 25 are the negative-bias questions, and questions 2, 4, 6, 12, and 14 are the positive-bias ones.

References

Agrawal, Kalin, and Batchelder, William H. ''Cultural Consensus Theory: Aggregating Signed Graphs under a Balance Constraint.'' Pp. 53–60 in *Social Computing, Behavioral-cultural Modeling, and Prediction* (SBP 2012), edited by J. Y. Shanchieh, A. M. Greenberg, and M. Endsley. Heidelberg, Germany: Springer.

Allum, Nick, Patrick Sturgis, Dimitra Tabourazi, and Ian Brunton-Smith. 2008. ''Science Knowledge and Attitudes across Cultures: A Meta-analysis.'' *Public Understanding of Science* 17(1):35–54.

American Association of Retired Persons. 1994. ''Images of Aging in America, 1994'' [Computer file]. icpsr03094-v2. Washington, DC: American Association of Retired Persons [producer], 1995. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2001.

Anders, Royce and Batchelder, William H. 2012. ‘‘Cultural Consensus Theory for Multiple Consensus Truths.’’ *Journal of Mathematical Psychology* 56(6):452–69.

Atran, Scott, Douglas Medin, Norbert Ross, Elizabeth Lynch, John Coley, Edilberto Ucan Ek, and Valentina Vapnarsky. 1999. ‘‘Folkecology and Commons Management in the Maya Lowlands.’’ *Proceedings of the National Academy of Sciences of the United States of America*, 96(13):7598–7603.

Baer, Roberta D., Susan C. Weller, Javier Garcia de Alba Garcia, Mark Glazer, Robert Trotter, Lee Pachter, and Robert E. Klein. 2003. ‘‘A Cross-cultural Approach to the Study of the Folk Illness Nervios.’’ *Culture, Medicine, and Psychiatry* 27(3):315–37.

Bann, Carla M., and Michael J. Schwerin. 2004. ‘‘Public Knowledge and Attitudes Scale Construction: Development of Short Forms’’ (Technical Report No. RTI Project No. 8796). Research Triangle Park, NC: RTI International.

Batchelder, William H., and Royce Anders. ‘‘Cultural Consensus Theory: Comparing Different Concepts of Cultural Truth.’’ *Journal of Mathematical Psychology* 56(5):316–32.

Batchelder, William H., Ece Kumbasar, and John P. Boyd. 1997. ‘‘Consensus Analysis of Three-way Social Network Data.’’ *Journal of Mathematical Sociology* 22(1):29–58.

Batchelder, William H., and A. Kimball Romney. 1988. ‘‘Test Theory without an Answer Key.’’ *Psychometrika* 53(1):71–92.

Bauer, Martin. 1996. ‘‘Socio-demographic Correlates of DK-responses in Knowledge Surveys: Self-attributed Ignorance of Science.’’ *Social Science Information Sur Les Sciences Sociales* 35(1):39–68.

Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise on the Sociology of Knowledge*. Garden City, NY: Anchor.

Carpini, Michael X. Delli, and Scott Keeter. 1993. ‘‘Measuring Political Knowledge—Putting First Things First.’’ *American Journal of Political Science* 37(4):1179–1206.

Clogg, Clifford C. 1982. ‘‘Using Association Models in Sociological Research: Some Examples.’’ *American Journal of Sociology* 88(1):114–34.

Clogg, Clifford C. 1984. ‘‘Some Statistical Models for Analyzing Why Surveys Disagree.’’ Pp. 319–66 in *Surveying Subjective Phenomena*, Vol. 2, edited by C. F. Turner and E. Martin. New York: Russell Sage.

Courtenay, Bradley C., and Craig Weidemann. 1985. ‘‘The Effects of a ‘Don’t Know’ Response on Palmore’s Facts on Aging Quizzes.’’ *The Gerontologist* 25(2):177–81.

De Boeck, Paul, and Mark Wilson. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

Durant, John R., Geoffrey A. Evans, and Geoffrey P. Thomas. 1989. ‘‘The Public Understanding of Science.’’ *Nature* 340(6228):11–14.

Feick, Lawrence F. 1989. ‘‘Latent Class Analysis of Survey Questions That Include Don’t Know Responses.’’ *Public Opinion Quarterly* 53(4):525–47.

Francis, Joe D., and Lawrence Busch. 1975. ‘‘What We Now Know about Don’t Knows.’’ *Public Opinion Quarterly* 39(2):207–18.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. New York: Chapman & Hall.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach*. 2nd ed. New York: Chapman & Hall.

Harris, Lori A., and Stephanie Dollinger. 2001. ‘‘Participation in a Course on Aging: Knowledge, Attitudes, and Anxiety about Aging in Oneself and Others.’’ *Educational Gerontology* 27(8):657–67.

Hayes, Bernadette C., and Vicki N. Tariq. 2000. ‘‘Gender Differences in Scientific Knowledge and Attitudes toward Science: A Comparative Study of Four Anglo-American Nations.’’ *Public Understanding of Science* 9(4):433–47.

Holtzman, Joseph M., and James D. Beck. 1979. ‘‘Palmore’s Facts on Aging Quiz—Reappraisal.’’ *The Gerontologist* 19(1):116–20.

Hopkins, Allison. 2011. ‘‘Use of Network Centrality Measures to Explain Individual Levels of Herbal Remedy Cultural Competence among the Yucatec Maya in Tabi, Mexico.’’ *Field Methods* 23(3):307–28.

Hruschka, Daniel J., Lynn M. Sibley, Nahid Kalim, and Joyce K. Edmonds. 2008. ‘‘When There Is More Than One Answer Key: Cultural Theories of Postpartum Hemorrhage in MATLAB, Bangladesh.’’ *Field Methods* 20(4):315–37.

Karabatsos, George, and William H. Batchelder. 2003. ‘‘Markov Chain Estimation Methods for Test Theory without an Answer Key.’’ *Psychometrika* 68:373–89.

Klemmack, David L. 1978. ‘‘Examination of Palmore Facts on Aging Quiz—Comment.’’ *The Gerontologist* 18(4):403–5.

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. 2002. ‘‘The Impact of ‘No Opinion’ Response Options on Data Quality—Non-attitude Reduction or an Invitation to Satisfice?’’ *Public Opinion Quarterly* 66(3):371–403.

Kruschke, John K. 2011. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. New York: Academic Press.

Liao, Tim Futing. 1995. ‘‘The Nonrandom Selection of Don’t Knows in Binary and Ordinal Responses: Corrections with the Bivariate Probit Model with Sample Selection.’’ *Quality and Quantity* 29(1):87–110.

Lizotte, Mary-Kate, and Andrew H. Sidman. 2009. ‘‘Explaining the Gender Gap in Political Knowledge.’’ *Politics and Gender* 5(2):127–51.

Luskin, Robert C., and John G. Bullock. 2011. ‘‘‘Don’t Know’ Means ‘Don’t Know’: DK Responses and the Public’s Level of Political Knowledge. *Journal of Politics* 73(2):547–57.

Macmillan, Neil A., and C. Douglas Creelman. 2005. *Detection Theory: A User’s Guide*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

Medin, Douglas L., Norbert O. Ross, Scott Atran, Douglas Cox, John Coley, Julia B. Proffitt, and Sergey Blok. 2006. ‘‘Folkbiology of Freshwater Fish.’’ *Cognition* 99(3):237–73.

Miller, Jon D. 1998. ‘‘The Measurement of Civic Scientific Literacy.’’ *Public Understanding of Science* 7(3):203–23.

Miller, Jon D. 2004. ‘‘Public Understanding of, and Attitudes toward, Scientific Research: What We Know and What We Need to Know.’’ *Public Understanding of Science* 13(3):273–94.

Miller, Jon D. 2010a. ‘‘Civic Scientific Literacy: The Role of the Media in the Electronic Era.’’ Pp. 44–63 in *Science and the Media*, edited by D. Kennedy and G. Overholser. Cambridge, MA: American Academy of Arts and Sciences.
Miller, Jon D. 2010b. ‘‘The Conceptualization and Measurement of Civic Scientific Literacy for the 21st Century.’’ Pp. 241–55 in *Science and the Educated American: A Core Component of Liberal Education*, edited by J. Meinwald and J. G. Hildebrand. Cambridge, MA: American Academy of Arts and Sciences.
Miller, Jon D. 2011. ‘‘The Sources and Impact of Civic Scientific Literacy.’’ Pp. 213–36 in *The Culture of Science: How the Public Relates to Science across the Globe*, edited by M. W. Bauer, R. Shukla, and N. Allum, New York: Routledge.
Miller Jon D., Eugenie C. Scott, and Shinji Okamoto. 2006. ‘‘Public Acceptance of Evolution.’’ *Science* 313(5788):765–66.
Miller, Richard B., and Richard A. Dodder. 1980. ‘‘A Revision of Palmore’s Facts on Aging Quiz.’’ *The Gerontologist* 20(6):673–79.
Mondak, Jeffrey J. 2001. ‘‘Developing Valid Knowledge Scales.’’ *American Journal of Political Science* 45(1):224–38.
Mondak, Jeffrey J., and Mary R. Anderson. 2004. ‘‘The Knowledge Gap: A Reexamination of Gender-based Differences in Political Knowledge.’’ *Journal of Politics* 66(2):492–512.
Mondak, Jeffrey J., and Damarys Canache. 2004. ‘‘Knowledge Variables in Cross-national Social Inquiry.’’ *Social Science Quarterly* 85(3):539–58.
Mondak, Jeffrey J., and Belinda Creel Davis. 2001. ‘‘Asked and Answered: Knowledge Levels When We Will Not Take ‘Don’t Know’ for an Answer.’’ *Political Behavior* 23(3):199–224.
National Science Board. 1996. ‘‘Science and Engineering Indicators 1996.’’ Technical report. Washington, DC: National Science Foundation.
National Science Board. 2006. ‘‘Science and Engineering Indicators 2006.’’ Technical report. Retrieved October 5, 2012 (http://www.nsf.gov/statistics/seind06/).
National Science Board. 2008. ‘‘Science and Engineering Iindicators 2008.’’ Technical report. Retrieved October 5, 2012 (http://www.nsf.gov/statistics/seind08/).
National Science Board. 2010. ‘‘Science and Engineering Indicators 2010.’’ Technical report. Retrieved September 19, 2012 (http://www.nsf.gov/statistics/seind10/).
National Science Board. 2012. ‘‘Science and Engineering Indicators 2012.’’ Technical report. Retrieved September 19, 2012 (http://www.nsf.gov/statistics/seind12/).
Norris, Joan E., Joseph A. Tindale, and Anne Martin Matthews. 1987. ‘‘The Factor Structure of the Facts on Aging Quiz.’’ *The Gerontologist* 27(5):673–76.
Oravecz, Zita, Joachim Vandekerckhove, and William H. Batchelder. Forthcoming. ‘‘Bayesian Cultural Consensus Theory.’’ *Field Methods*.
Palmore, Erdman. 1977. ‘‘Facts on Aging—Short Quiz.’’ *The Gerontologist* 17(4):315–20.
Palmore, Erdman. 1978. ‘‘Examination of Palmore Facts on Aging Quiz—Reply.’’ *The Gerontologist* 18(4):405–6.
Palmore, Erdman. 1980. ‘‘The Facts on Aging Quiz—A Review of Findings.’’ *The Gerontologist* 20(6):669–72.
Palmore, Erdman. 1982. ‘‘Attitudes toward the Aged—What We Know and Need to Know.’’ *Research on Aging* 4(3):333–48.
Palmore, Erdman. 1998. *Facts on Aging Quiz*. 2nd ed. New York: Springer.

Palmore, Erdman. 2005. ''Three Decades of Research on Ageism.'' *Generations* 29(3):87–90.

Pardo, Rafael, and Félix Calvo. 2004. ''The Cognitive Dimension of Public Perceptions of Science: Methodological Issues.'' *Public Understanding of Science* 13(3):203–27.

Plummer, Martyn. 2003. ''JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.'' Pp. 20–22 in *Proceedings of the Third International Workshop on Distributed Statistical Computing*, edited by F. Leisch, K. Hornik, and A. Zeileis.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Romney, A. Kimball, Susan C. Weller, and William H. Batchelder. 1999. ''Cultural Consensus Theory.'' Pp. 208–209 in *The MIT Encyclopedia of the Cognitive Sciences*, edited by R. Wilson and F. Keil. Cambridge, MA: MIT Press.

Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. ''Culture as Consensus: A Theory of Culture and Informant Accuracy.'' *American Anthropologist* 88(2):313–38.

Seufert, Robert L., and Mark A. Carrozza. 2002. ''A Test of Palmore's Facts on Aging Quizzes as Alternative Measures.'' *Journal of Aging Studies* 16:279–94.

Snijders, Tom A. B., and Roel J. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Spearman, Charles E. 1904. '''General Intelligence' Objectively Determined and Measured.'' *American Journal of Psychology* 15(1):72–101.

Stuart-Hamilton, Ian, and Berenice Mahoney. 2003. ''The Effect of Aging Awareness Training on Knowledge of, and Attitudes towards, Older Adults.'' *Educational Gerontology* 29(3):251–60.

Sturgis, Patrick, Nick Allum, and Patten Smith. 2008. ''An Experiment on the Measurement of Political Knowledge n Surveys.'' *Public Opinion Quarterly* 72(1):90–102.

Tsai, Tsung-han, and Jeff Gill. 2012. ''superdiag: A Comprehensive Test Suite for Markov Chain Non-convergence.'' *The Political Methodologist* 19(2):12–18.

Waubert de Puiseau, Berenike, André Aßfalg, Edgar Erdfelder, and Daniel M. Berstein. 2012. ''Extracting the Truth from Conflicting Eyewitness Reports: A Formal Modeling Approach.'' *Journal of Experimental Psychology: Applied* 18(4):390–403.

Weller, Susan C. 1984a. ''Consistency and Consensus among Informants: Disease Concepts in a Rural Mexican Village.'' *American Anthropologist* 86(4):966–75.

Weller, Susan C. 1984b. ''Cross-cultural Concepts of Illness—Variation and Validation.'' *American Anthropologist* 86(2):341–51.

Weller, Susan C. 2007. ''Cultural Consensus Theory: Applications and Frequently Asked Questions.'' *Field Methods* 19(4):339–68.

Weller, Susan C., Roberta D. Baer, Javier Garcia de Alba Garcia, and Ana L. Salcedo Rocha. 2012. ''Explanatory Models of Diabetes in the U.S. and Mexico: The Patient-Provider Gap and Cultural Competence.'' *Social Science and Medicine* 75(6):1088–96.

Weller, Susan C., Roberta D. Baer, Lee M. Pachter, Robert T. Trotter, Mark Glazer, Javier Garcia de Alba Garcia, and Robert Klein. 1999. ''Latino Beliefs about Diabetes.'' *Diabetes Care* 22(5):722–28.

Weller, Susan C., Lee M. Pachter, Robert T. Trotter, Roberta D. Baer. 1993. ‘‘Empacho in Four Latino Groups: A Study of Intra- and Inter-cultural Variation in Beliefs.’’ *Medical Anthropology* 15(2):109–36.

World Values Survey Association. 2009. ‘‘World Values Survey 1981–2008 Official Aggregate v.20090902’’ (http://www.worldvaluessurvey.org).

York-Crowe, Emily E., Marney A. White, Sahasporn Paeratakul, and Donald A. Williamson. 2006. ‘‘The Diet and Health Knowledge Survey: Development of a Short Interview Format.’’ *Eating Behaviors* 7(3):235–42.

Yoshino, Ryozo. 1989. ‘‘An Extension of the ‘Test Theory without Answer Key’ by Batchelder and Romney and Its Application to an Analysis of Data on National Consciousness’’ (in Japanese). *Proceedings of the Institute of Statististical Mathematics* 37:171–88.

Zigerell, L. J. 2012. ‘‘Science Knowledge and Biblical Literalism.’’ *Public Understanding of Science* 21(3):314–22.

Author Biographies

Zita Oravecz earned her doctorate in quantitative psychology at the University of Leuven, Belgium. Her thesis focused on the dynamics of time-evolving psychological processes and interindividual differences therein. She continued investigating between-person variability in decision processes as a postdoctoral researcher at the University of California, Irvine. Her research interests involve hierarchical Bayesian modeling, psychometrics, intensive longitudinal data, and latent variable modeling. She also developed user-friendly software applications that carry out parameter estimation for complex hierarchical process models.

Katherine Faust is a professor of sociology and a member of the Institute for Mathematical Behavioral Sciences at the University of California, Irvine. Her current research focuses on comparing network structural signatures across different forms of social relations and animal species; developing methodology for complex network structures, including asymmetry in multirelational networks, and constraints on local network properties; and understanding relationships between social networks and demographic processes. She is coauthor (with Stanley Wasserman) of the book *Social Network Analysis: Methods and Applications* (Cambridge University Press) and numerous articles on social network methodology.

William H. Batchelder is a professor of cognitive sciences and a member and former director of the Institute for Mathematical Behavioral Sciences at the University of California, Irvine. His research interests involve methodology, measurement, and modeling in the social and behavioral sciences, especially cultural anthropology, social networks, and cognitive psychology. He is a former editor of the *Journal of Mathematical Psychology* and an elected member of the Society of Experimental Psychologists. He has also achieved the title of Senior Master in the United States Chess Federation.