Predicting judgments of food healthiness with deep latent-construct cultural consensus theory

Necdet Gürkan (ngurkan@stevens.edu)
School of Business, Stevens Institute of Technology
Hoboken, NJ 07030, USA

Jordan W. Suchow (jws@stevens.edu)
School of Business, Stevens Institute of Technology
Hoboken, NJ 07030, USA

Abstract

Deep neural network representations of entities can serve as inputs to computational models of human mental representations to predict people's behavioral and physiological responses to those entities. Though increasingly successful in their predictive capabilities, the implicit notion of "human" that they rely upon often glosses over individual-level differences in beliefs, attitudes, and associations, as well as group-level cultural constructs. In this paper, we model shared representations of food healthiness by aligning learned word representations with the consensus among a group of respondents. To do so, we extend Cultural Consensus Theory to include latent constructs structured as fine-tuned word representations. We then apply the model to a dataset of people's judgments of food healthiness. We show that our method creates a robust mapping between learned word representations and culturally constructed representations that guide consumer behavior.

Keywords: cultural consensus theory, individual differences, bayesian modeling, food beliefs

Introduction

People form subjective evaluations about what constitutes a healthy diet (Ford, Bergmann, Boeing, Li, & Capewell, 2012; Capewell & O'Flaherty, 2011; Beaglehole et al., 2011) through their beliefs, associations, attitudes, and knowledge about food healthiness (Messer, 1984). The diverse psychological, cognitive, and social elements that impact consumer choices make it impossible to identify a single correct answer for what makes food appear healthy (Lobstein & Davies, 2009). Consequently, understanding individuals' health beliefs and perspectives on healthy food is crucial for developing culturally and socially tailored behavior change interventions (Barrera Jr, Castro, Strycker, & Toobert, 2013; Sassi, Cecchini, Lauer, & Chisholm, 2009).

Healthy food choice is a complex behavior encompassing cultural (e.g., customs, social norms), psychological (e.g., body image), and social factors (e.g., price availability, ethical concerns) (Pearcey & Zhan, 2018; Arganini, Saba, Comitato, Virgili, & Turrini, 2012). Previous research examining healthy food perception in diverse groups has revealed the distinct ways individuals conceptualize healthy food (Banna, Gilliland, Keefe, & Zheng, 2016). Cross-cultural research suggests that some cultures have similar interpretations of healthy food choice and nutritional intake (Akamatsu, Maeda, Hagihara, & Shirakawa, 2005), while others differ (Banna et al., 2016). Gender and level of educational attainment also play a role in judgments of what makes food healthy (Arganini et al., 2012; Turrell & Kavanagh, 2006). In addition to idiosyncratic features, front-of-package labeling is argued to be an influential factor on people's judgments of food healthiness (Orquin, 2014). However, healthiness judgments derived from front-of-package labeling are also influenced by idiosyncratic features (Schuldt, 2013).

Recent machine learning methods, capable of learning expressive feature representations of high-dimensional concepts, have been employed to model social, factual, probability, and semantic judgments (Bhatia, 2017; Nousi & Tefas, 2017; Poeppel, 2012). For instance,Bhatia (2019) developed an accurate mapping between machine-generated word representations (i.e., word embeddings) and human risk perception. In a related project, word representations were used to predict people's judgments of food healthiness and uncover psychological associations between healthy and unhealthy foods (Gandhi, Zou, Meyer, Bhatia, & Walasek, 2022). These same methods can be applied to any human judgment that relies on semantic knowledge captured by a machine-derived vector space model of words.

Although models that align machine semantic representations with human judgments are increasingly successful in predicting behavioral responses, the implicit notion of "human" upon which they rely often overlooks individual-level differences in subjective beliefs, attitudes, and associations, as well as variations derived from group-level cultural constructs. For instance, one subculture may perceive flavored yogurt as unhealthy, while another may consider it an integral component of their diet. As a result, different subsets of the population may hold quite distinct representations of flavored yogurt when judging food healthiness. Modeling both individual-level and group consensus-level variation enhances the alignment of machine representations with human psychological inferences (Gurkan & Suchow, 2022a).

When making subjective judgments of food healthiness, groups of respondents may share specific knowledge, beliefs, or preferences that are unknown a priori to researchers. Cultural Consensus Theory (CCT) is a statistical framework that can be employed to infer the cultural beliefs influencing social practices and the degree to which individuals know or express those beliefs ((Romney, Weller, & Batchelder, 1986). These models provide an opportunity to study individual differences in whether a group member conforms to the consensus in a community, and they allow people to differ in both

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

their level of cultural knowledge and response biases. Researchers have applied the CCT framework to find practical and concise definitions of beliefs that are accepted by a group sharing common knowledge. These models have been widely used to study mental health (van den Bergh et al., 2020), cognitive evaluation (Heshmati et al., 2019), eyewitness testimony (Waubert de Puiseau, Aßfalg, Erdfelder, & Bernstein, 2012)), (Rinne & Fairweather, 2012), and online communities (Gurkan & Suchow, 2022b).

In this study, we align machine word-vector representations with the consensus among a group of respondents by extending CCT to include latent constructs structured as vector representations of words. CCT is a model-based statistical technique used to estimate the consensus among a group of respondents when the ground truth is unavailable or undefined (e.g., in collective memories of events, beauty norms, or social relationships in covert networks). CCT offers a practical basis for capturing individual and group differences using participant-level data. However, the base model has several limitations, such as neglecting the correlation between items and lacking a mechanism to map from features of the stimulus or question to resultant answers and cultural consensuses. In this paper, we extend CCT by enabling culturally held beliefs to take the form of a latent construct, a regression model that maps from a question (represented as a vector) to a consensus answer. Our approach aligns learned machine representations with group consensus and individual levels, thereby capturing variation in psychological processes and behaviors across individuals and groups.

The structure of this paper is as follows. We begin by reviewing the CCT model for continuous responses. Next, we introduce our extended latent-construct CCT model. Following that, we describe the dataset of food healthiness judgments. Lastly, we present and discuss the results of fitting the model.

Cultural consensus theory

Cultural Consensus Theory (CCT) is a mathematical framework that measures each respondent's cultural knowledge while estimating the culturally "correct" answers to a series of questions defined by group beliefs or norms. The CCT model jointly estimates (1) an individual's level of cultural knowledge from their agreement with a consensus truth and (2) the consensus truth itself from a weighted average of responses, giving higher weight to individuals with higher cultural competence. The first CCT model, the General Condorcet Model (GCM), was developed for binary data (true/false responses) and assumes that the consensus truth of each item is also a binary value (Romney et al., 1986). The GCM has been widely used in the social and behavioral sciences (Weller, 2007).

Batchelder and Anders (2012) introduced an alternate assumption to the GCM to extend it to continuous truths. An extensive CCT model for ordinal data was developed using a Gaussian appraisal model (Anders & Batchelder, 2015). In addition, CCT models for continuous response data were developed to estimate and detect cultural consensuses, informant knowledge, response biases, and item difficulty from continuous data (Anders, Oravecz, & Batchelder, 2014; Batchelder & Romney, 1988; Batchelder, Strashny, & Romney, 2010).

Here, we describe a Continuous Response Model (CRM), developed by Anders et al. (2014) that allows for multiple consensus truths, which serves as the basis for our extension to the model.

Continuous cultural consensus theory

As a starting point, consider the Continuous Response Model (CRM) (Anders et al., 2014), a cultural consensus model for continuous data derived from observations of the random response profile matrix $\mathbf{X}_{ik} = (X_{ik})_{N\times M}$ for N respondents and M items, where each respondent's response falls within $(0,1)$ or a finite range that permits a linear transformation to $(0,1)$. The CRM links the random response variables in (0, 1) to the real line with the logit transform, $X^* = \text{logit}(X_{ik})$. Therefore, each item also has a consensus value in $(-\infty,\infty)$.

The CRM is formalized and further explained by the following axioms:

Axiom 1 (*Cultural truths*). There is a collection of of $V \geq 1$ latent cultural truths, $\{T_1,...,T_v,...,T_V\}$, where $T_V \in \prod_{k=1}^{M}(-\infty,\infty)$. Each participant, i, responds according to only one cultural truth (set of consensus locations), as T_{Ω_i}, where $\Omega_i \in \{1,....,V\}$, and parameter $\Omega = (\Omega_i)_{1\times N}$ denotes the cultural membership for each informant.

Axiom 2 (*Latent Appraisals*). It is assumed that each participant draws a latent appraisal, Y_{ik}, of each $T_{\Omega_{ik}}$, in which $Y_{ik} = T_{\Omega_i k} + \varepsilon_{ik}$, The ε_{ik} error variables are distributed normal with mean 0 and standard deviation σ_{ik}.

Axiom 3 (*Conditional Independence*). The ε_{ik} are mutually stochastically independent.

Axiom 4 (*Precision.*). There are knowledge competency parameters $\mathbf{E} = (E_i)_{1\times N}$ with all $E_i > 0$, and item difficulty parameters specific to each cultural truth $\Lambda = (\lambda_k)_{1\times M}$, $\lambda_k > 0$ such that

$$\sigma_{ik} = \lambda_k / E_i. \tag{1}$$

If all item difficulties are equal, then each λ_k is set to 1.

Axiom 5 (*Response Bias*). There are two respondent bias parameters that act on each respondent's latent appraisals, Y_{ik}, to arrive at the observed responses, the X_{ik}. These include a scaling bias, $\mathbf{A} = (a_i)_{1\times N}, a_i > 0$; and shifting bias $\mathbf{B} = (b_i)_{1\times N}, -\infty < b_i < \infty$, where

$$X^*_{ik} = a_i Y_{ik} + b_i. \tag{2}$$

These axioms are developed to model the continuous response of respondents that differ in cultural competency, E_i, and response biases, a_i and b_i, to items that have different shared latent truth values. The respondents have a latent appraisal of these item values with a mean at the item's consensus location plus error, which depends on their competence level and the item difficulty. Axiom 1 locates the item truth values in the continuum. Axiom 2 defines the appraisal error is normally distributed with mean zero. Axiom 3 sets the appraisals to be conditionally independent given the respondents' cultural truth and the error standard deviations. Axiom 4 specifies the standard appraisal error that depends on the respondent's competence and item difficulty. Axiom 5 sets each respondent's response shift and scale biases.

Extending CCT with infinite word representation latent constructs

CCT operationalizes the structure of culturally held beliefs as a lookup table, with keys that are questions and values that are answers. The questions and answers themselves have no defined internal structure and are not linked to each other in any way, except through correlations across respondents' answers. However, such a formulation has several limitations. First, because it treats each question/answer pair independently, information gleaned from one question does not inform our understanding of other questions. Second, for the same reason, the number of questions that must be tested to characterize a culture scales linearly with the number of culturally held beliefs. Third, there is no way to leverage insights from existing knowledge bases that provide structured information about known entities and their relations.

In this study, we propose to extend CCT by enabling culturally held beliefs to take the form of mathematical objects more complex than a lookup table. We operationalize culturally held beliefs as an algorithmic latent construct, a function that maps from a question to a consensus answer through an intermediate representation. Specifically, we consider an intermediate representation that has the structure of a learned word representation, which is fine-tuned with a linear readout layer specific to the culture.

Extending CCT in this way enables us to create a mapping between word embeddings, ϕ_k, and the cultural consensus, T_{vk}. To fit the latent construct, we introduce a latent variable, ω_{Ω_i}, that represents the regression weights for each participant's cultural membership, Ω_i. The relation between the weights of latent construct, ω_{Ω_i}, and high-dimensional expressive machine features, ϕ_k, is given by the regression equation

$$\begin{aligned} T_{vk} &= \phi_k \omega_{\Omega_i}^T \\ Y_{ik} &= T_{vk} + \varepsilon_{ik}, \end{aligned} \tag{3}$$

where ε_{ik} is the error variable in Axiom 4 (Eq. 1). We replace the consensus location described in Axiom 1 with a function that takes as input the machine features and corresponding weights for each feature.

In this work, we use Bayesian Ridge regression to regularize the weights in the latent construct. The prior for the coefficients, ω_{Ω_i}, is given by a spherical Gaussian:

$$p(\omega_{\Omega_i} \mid \zeta) = \text{Normal}(\omega_{\Omega_i} \mid 0, \zeta^{-1}\mathbf{I}_p), \tag{4}$$

with the prior over ζ assumed to be Gamma distributed, the conjugate prior for the precision of the Gaussian.

When there are multiple cultures, CCT analyzes eigenvalues obtained from the cross-participant correlation matrix to determine the number of cultures present. Two key problems with this approach are that it assumes a finite-dimensional representation that correctly characterizes the features of observed data and that there are few missing values in the observed data. We further extended the CCT model to enable an unbounded number of cultures using a Dirichlet Process prior (via a stick-breaking process) over the cluster assignments. This modification produces a Bayesian nonparametric model, where the number of instantiated cultures grows with the complexity of the observed data. As opposed to the original CCT framework, which uses a fixed number of cultures tuned by the experimenter, our Bayesian nonparametric variant of the model provides a posterior over the entire space of partitions.

Hierarchical specification of the extended CCT

In this section, we specify the extended CCT hierarchically (Lee, 2011), where population distributions are specified for the parameters using hyperparameters. These hyperparameters are estimated from their own distributions and can represent the central tendency within each trait across items or participants, which may be unique to each dataset. The hierarchical structure of our generative model is as follows.

$$\begin{aligned}
\omega_{\Omega_i} &\sim \text{Normal}(0, \zeta^{-1}) && \text{Coefficient weights} \\
T_{vk} &= \phi_k \omega_{\Omega_i}^T && \text{Latent construct item location} \\
\log(E_i) &\sim \text{Normal}(\alpha_{E_{\Omega_i}}, \kappa_{E_{\Omega_i}}) && \text{Informant competency} \\
\log(a_i) &\sim \text{Normal}(\mu_{a_i}, \tau_{a_i}) && \text{Informant scaling bias} \\
b_i &\sim \text{Normal}(\mu_{b_i}, \tau_{b_i}) && \text{Informant shifting bias} \\
\Omega_i &\sim \text{Categorical}(\pi) && \text{Group membership} \\
\pi &\sim \text{stickbreaking}(\beta) && \text{Pr. of group membership} \\
\beta &\sim \text{Beta}(1, \delta) && \text{Group sparsity}
\end{aligned}$$

We used Bayesian Ridge Regression as a latent construct to generate the item consensus, T_{vk}. The location of the item consensus is calculated by taking the dot product of two row matrices. The other model parameters, E_i, and a_i, which are each located on the positive half-line, are log-transformed to the real line and also assumed to be sampled from a normal population-level distribution. Item difficulty, λ_i, is set to 1, under the assumption that each item is equally difficult to rate. The shift bias, b_i, is located on the real line, paramaterized with a mean and precision (inverse variance). Note that the informants' competence parameter remains singly-indexed by i, through an indexing technique in which their distribution is specified by their group membership Ω_i. Culture assignments are derived via a stick-breaking prior, and this allows for varying probabilities of being in any of V groups; note that V is unknown priori and needs to be estimated from observed data. The sparsity of cultures, β, is sampled from a beta distribution. The variables Ω_i and π are removed for the single-truth variant of our model. In this model, we assume that the participant's shift biases are relatively small, which has the benefit of improving convergence of the MCMC sampler.

The hyperparamaters are set as follows:

$$\begin{aligned} \zeta_{T_v} &\sim \text{Gamma}(1, 0.1) & \delta &\sim \text{Gamma}(1,1) \\ \alpha_{E_{\Omega_i}} &\sim \text{Gamma}(5,1) & \kappa_{E_{\Omega_i}} &\sim \text{Gamma}(1,1) \\ \mu_{a_i} &\sim \text{Gamma}(3,1) & \tau_{a_i} &\sim \text{Gamma}(1,1) \\ \mu_{b_i} &= 0 & \tau_{b_i} &\sim \text{Gamma}(1,50) \end{aligned}$$

The concentration parameter, δ, controls the prior over the number of clusters; a large concentration parameter leads to a greater number of clusters.

Method

Data

We applied our extended CCT model to a dataset of people's judgments of food healthiness (Gandhi et al., 2022). The dataset contains judgments of 149 participants on a diverse set of 172 foods. Participants were asked to judge the healthiness of food on a scale ranging from -100 (extremely unhealthy) to +100 (extremely healthy). We rescaled ratings to fall in the interval [0, 1] for our extended CCT model. We used word2vec word representations (Mikolov, Chen, Corrado, & Dean, 2013) to obtain an embedding for each food. These word vectors place 3 million English words (and phrases) in a 300-dimensional space where the location of words is derived from the contexts in which the words are found in a large corpus, and the distance between words captures something about their semantic relatedness. We used the ratings of 138 food items for training and the remainder (34 food items) for validation, an 80-20 split. Here, the aim is to predict individual ratings of healthiness for the food items in the validation set using only the food items in the training set.

Implementation: The model was implemented in NumPyro (Phan, Pradhan, & Jankowiak, 2019) with the JAX backend (Bradbury et al., 2020). The model components were integrated into a single likelihood function and a set of prior distributions, needed to infer a posterior over the unobserved variables in our model using the Gibbs Sampler (Liu, 1996) combined with No-U-Turn Sampler (NUTS) (Hoffman, Gelman, et al., 2014), a standard Markov chain Monte Carlo sampling algorithm, as implemented in Numpyro. We used 2 chains with 10,000 warm up samples and 10,000 draw samples, thereby obtaining 20,000 posterior samples. We ensured that the posterior had converged by ensuring there were not divergence transitions.

Results

In this section, we demonstrate the predictive accuracy of our extended CCT model with respect to people's subjective judgments of food healthiness. To determine whether the participant-level data provides evidence of one vs. multiple consensuses, we separately fit a single-truth version of the extended CCT, which assumes that there is only one consensus for a given item, and a multiple-culture version of the extended CCT, as described above. As shown in Fig. 1, predictive accuracy is improved when we create a culturally specific mapping between learned word representations and consensus beliefs. Our extended multiple-culture truths CCT model has a lower root mean square error on the held-out food items (RMSE = .20) than the single-truth CCT model (RMSE = .24). Also, the multiple-culture truths CCT model explains a greater proportion of the variance in the predicted values ($R^2 = .48$), compared to the single-culture truth CCT model ($R^2 = .32$). Additionally, the extended multiple-culture model includes the single-culture model as a special case; thus, our finding of an assignment of respondents to multiple cultures is another source of evidence that a multiple-culture model fits the data better.

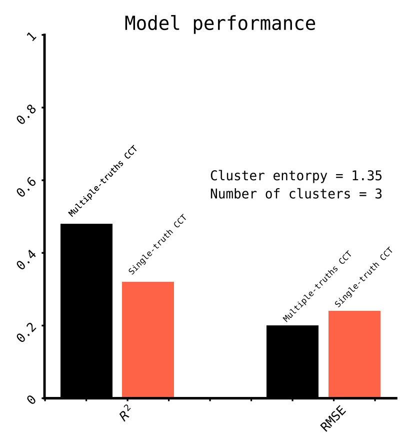

Figure 1: The black and red bars represent the extended multiple-culture truths and single-truth CCT models, respectively. **Left**: Performance (R^2) for the extended single-truth (red bar) and multiple-truths (black bar) CCT models. **Right:** Prediction accuracy (RMSE) using the extended single-truth (red bar) and multiple-culture truths (black bar) CCT models.

Fig. 1 also reports the best-fit number of instantiated cultures. While the number of cultures indicates the number of consensuses in the data, it does not inform us about the uniformity of the cultural assignment distribution. An entropy-based metric addresses this problem by estimating the uncertainty of the cultural allocation of an unknown randomly chosen data point given a particular distribution of culture assignments. In Fig. 1, the smaller value of entropy shows that there are a few large clusters, and the larger values of entropy are associated with more evenly distributed cultures. In the analysis, the final cultural assignment is determined by the modal membership across the final 100 posterior samples. Fourteen respondents were clustered into culture 1, 73 were clustered into culture 2, and 62 were clustered into culture 3.

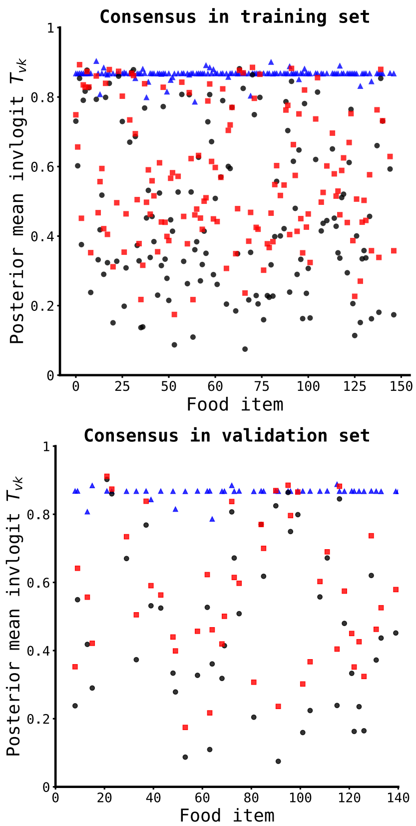

Figure 2: **Top**: Estimated consensus values for each inferred cultures in the training set. **Buttom:** Predicted consensus values using learned regression weights for each inferred cultures in the validation set.

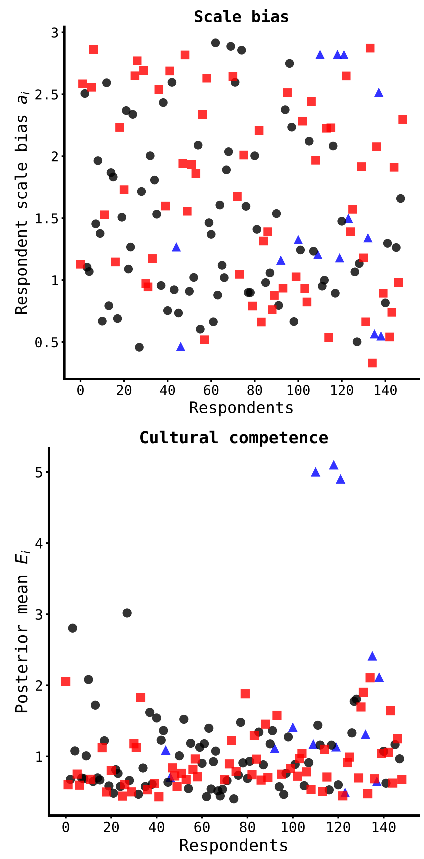

Figure 3: Individual scale response biases, a_i plotted for each informant (top). Respondent competencies, E_i, (bottom). The blue triangle, red square, and black circles, respectively denote the three separate cultures' informant membership.

Fig. 2 shows the posterior mean inverse logit item values for each of the three cultures recovered and predicted. The respondents assigned to the blue triangle culture appear to consistently rate the food items as being healthier than the other cultures do. This may signal that these participants did not provide quality data, by rating every food item high, or perhaps they are a group of participants who genuinely believe that every food item can be healthy in the right context. For example, the term "all foods fit" refers to the idea that there are no "good" or "bad" foods, which emphasizes that every food can be beneficial in moderate portions (May, Galper, & Carr, 2004).

We next present details of the learned model. As shown in Figure 2, the red square and black circle cultures have similar beliefs about some of the food items. Individuals from these cultures have similar beliefs on the healthiness of honeydew melon and aloe vera juice. These two cultures seem to disagree the most on the healthiness of macaroons. The healthiest-seeming foods among the three cultures are apples, bananas, and kale, which align with the general perception of participants in this study. Although the blue triangle culture gave high ratings to the food items, donuts are found to be the least healthy food among all for the blue triangle culture. The black circle and red square cultures perceive frozen yogurt and cola as the unhealthiest food items.

It is important to estimate culturally tailored beliefs about what constitutes healthy food when promoting healthy food interventions. Our extended CCT model is capable of predicting culturally held beliefs about the food healthiness for food items outside the training set. The extended CCT takes advantage of the semantic correlation of deep representations of food items. The extended CCT predicts that artichoke and corn flakes are perceived as unhealthy choices by the black circle and red square cultures, respectively. Although these two cultures view corn flakes and artichoke as unhealthy food items, the blue triangle culture perceives them as healthy or adopts an "all foods fit" perspective. The most controversial food item among the cultures is spaghetti.

The figure 3 contains the standard appraisal errors, E_i, of each informant. Clusters have comparable average standard appraisal errors at $E_v = [2.5, 1.4, 1.1]$. The level of competency of these differently grouped informants show how well their beliefs and attitudes towards food healthiness fit into the cluster they are assigned to. For example, the three participants from the blue triangle cluster demonstrate a strong cultural membership. The figure 3 also provides the individual response bias behaviors of the informant, with posterior a_i and their cultural membership, from cluster mode.

Discussion

In this paper, we extended the CCT by enabling culturally held beliefs to take the form of a latent construct that maps from food items to a consensus judgment of food healthiness through a word vector representation fine-tuned using Bayesian Ridge regression. The main aim is to gather opinions from a diverse group of people, including experts and non-experts, in order to determine shared beliefs while also pinpointing unique and group-specific cultural differences. By adding elements from deep neural networks to the CCT, we can evaluate cultural consensus for any subject using preexisting networks or other accessible embeddings. Our results show that taking into account variations in individual and group-level consensus improves the alignment of word representations with people's judgments of food healthiness.

We note that our model is not restricted to people's judgments of food healthiness. Integrating features extracted from deep neural networks into the CCT provides a unique opportunity to estimate a cultural consensus from any kind of entity for which a pretrained network or other embedding is available. Additionally, machine-learning techniques have made it possible to extract rich quantitative representations from other modalities, such as acoustic and visual. These vector-space representations can also be aligned with people's subjective judgments and psychological inferences using the extended CCT model to form models of consensus beliefs in other domains.

The method and associated developments can have broad scientific applicability in (1) domains where social scientists currently study cultural variation but do not leverage Artificial Intelligence (AI) and machine-learning technologies, and (2) domains where computer science and behavioral scholars apply AI technologies to model human behavior without considering cultural variation. Further, the proposed deep CCT model has several benefits over the existing state of the art in terms of modeling a cultural consensus. First, consensus judgments among a group can be estimated with respect to objects, events, or concepts absent from the training set by encoding them into the network and applying the learned regression weights. Second, judgments can leverage state-of-the-art entity and event representations from the machine learning literature. And third, information is not siloed across entities and events: what we learn from judgments of one question informs judgments of other events, concepts, and objects.

References

Akamatsu, R., Maeda, Y., Hagihara, A., & Shirakawa, T. (2005). Interpretations and attitudes toward healthy eating among japanese workers. *Appetite*, *44*(1), 123–129.

Anders, R., & Batchelder, W. H. (2015). Cultural consensus theory for the ordinal data case. *Psychometrika*, *80*(1), 151–181.

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

Arganini, C., Saba, A., Comitato, R., Virgili, F., & Turrini, A. (2012). Gender differences in food choice and dietary intake in modern western societies. *Public Health-Social and Behavioral Health*, *4*, 83–102.

Banna, J. C., Gilliland, B., Keefe, M., & Zheng, D. (2016). Cross-cultural comparison of perspectives on healthy eating among chinese and american undergraduate students. *BMC Public Health*, *16*(1), 1–12.

Barrera Jr, M., Castro, F. G., Strycker, L. A., & Toobert, D. J. (2013). Cultural adaptations of behavioral health interventions: a progress report. *Journal of Consulting and Clinical Psychology*, *81*(2), 196.

Batchelder, W. H., & Anders, R. (2012). Cultural consensus theory: Comparing different concepts of cultural truth. *Journal of Mathematical Psychology*, *56*(5), 316–332.

Batchelder, W. H., & Romney, A. K. (1988). Test theory without an answer key. *Psychometrika*, *53*(1), 71–92.

Batchelder, W. H., Strashny, A., & Romney, A. K. (2010). Cultural consensus theory: Aggregating continuous responses in a finite interval. In *International Conference on Social Computing, Behavioral Modeling, and Prediction* (pp. 98–107).

Beaglehole, R., Bonita, R., Horton, R., Adams, C., Alleyne, G., Asaria, P., ... others (2011). Priority actions for the non-communicable disease crisis. *The Lancet*, *377*(9775), 1438–1447.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1.

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, *65*(8), 3800–3823.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., & Wanderman-Milne, S. (2020). JAX: composable transformations of Python+ NumPy programs, 2018. *URL http://github. com/google/jax*, *4*, 16.

Capewell, S., & O'Flaherty, M. (2011). Rapid mortality falls after risk-factor changes in populations. *The Lancet*, *378*(9793), 752–753.

Ford, E. S., Bergmann, M. M., Boeing, H., Li, C., & Capewell, S. (2012). Healthy lifestyle behaviors and all-cause mortality among adults in the united states. *Preventive Medicine*, *55*(1), 23–27.

Gandhi, N., Zou, W., Meyer, C., Bhatia, S., & Walasek, L. (2022). Computational methods for predicting and understanding food judgment. *Psychological Science*, *33*(4), 579–594.

Gurkan, N., & Suchow, J. (2022a). Cultural alignment of machine-vision representations. In *SVRHM Workshop, NeurIPS*.

Gurkan, N., & Suchow, J. W. (2022b). Learning and enforcing a cultural consensus in online communities. In *Pro-*

ceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44).

Heshmati, S., Oravecz, Z., Pressman, S., Batchelder, W. H., Muth, C., & Vandekerckhove, J. (2019). What does it mean to feel loved: Cultural consensus and individual differences in felt love. *Journal of Social and Personal Relationships*, *36*(1), 214–243.

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7.

Liu, J. S. (1996). Peskun's theorem and a modified discrete-state gibbs sampler. *Biometrika*, *83*(3).

Lobstein, T., & Davies, S. (2009). Defining and labelling 'healthy'and 'unhealthy'food. *Public Health Nutrition*, *12*(3), 331–340.

May, M., Galper, L., & Carr, J. H. (2004). *Am i hungry: What to do when diets don't work*. Am I Hungry.

Messer, E. (1984). Anthropological perspectives on diet. *Annual Review of Anthropology*, *13*(1), 205–249.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nousi, P., & Tefas, A. (2017). Discriminatively trained autoencoders for fast and accurate face recognition. In *International conference on engineering applications of neural networks* (pp. 205–215).

Orquin, J. L. (2014). A brunswik lens model of consumer health judgments of packaged foods. *Journal of Consumer Behaviour*, *13*(4), 270–281.

Pearcey, S. M., & Zhan, G. Q. (2018). A comparative study of american and chinese college students' motives for food choice. *Appetite*, *123*, 325–333.

Phan, D., Pradhan, N., & Jankowiak, M. (2019). Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv preprint arXiv:1912.11554*.

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, *29*(1-2), 34–55.

Rinne, T., & Fairweather, J. (2012). A mixed methods approach: Using cultural modeling and consensus analysis to better understand new zealand's international innovation performance. *Journal of Mixed Methods Research*, *6*(3), 166-183.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*(2), 313–338.

Sassi, F., Cecchini, M., Lauer, J., & Chisholm, D. (2009, 01). Improving lifestyles, tackling obesity: The health and economic impact of prevention strategies. *OECD, Directorate for Employment, Labour and Social Affairs, OECD Health Working Papers*. doi: 10.1787/220087432153

Schuldt, J. P. (2013). Does green mean healthy? nutrition label color affects perceptions of healthfulness. *Health Communication*, *28*(8), 814–821.

Turrell, G., & Kavanagh, A. M. (2006). Socio-economic pathways to diet: modelling the association between socio-economic position and food purchasing behaviour. *Public Health Nutrition*, *9*(3), 375–383.

van den Bergh, D., Bogaerts, S., Spreen, M., Flohr, R., Vandekerckhove, J., Batchelder, W. H., & Wagenmakers, E.-J. (2020). Cultural consensus theory for the evaluation of patients' mental health scores in forensic psychiatric hospitals. *Journal of Mathematical Psychology*, *98*, 102383.

Waubert de Puiseau, B., Aßfalg, A., Erdfelder, E., & Bernstein, D. M. (2012). Extracting the truth from conflicting eyewitness reports: A formal modeling approach. *Journal of Experimental Psychology: Applied*, *18*(4), 390.

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field methods*, *19*(4), 339–368.