Predictive evaluation of human value segmentations*

Kristoffer Jon Albers[1], Morten Mørup[1], Mikkel N. Schmidt[1],
Fumiko Kano Glückstad[2]

Abstract

Data-driven segmentation is an important tool for analysing patterns of associations in social survey data; however, it remains a challenge to compare the quality of segmentations obtained by different methods. We present a statistical framework for quantifying the quality of segmentations of human values, by evaluating their ability to predict held-out data. By comparing clusterings of human values survey data from the forth round of European Social Study (ESS-4), we show that demographic markers such as age or country predict better than random, yet are outperformed by data-driven segmentation methods. We show that a Bayesian version of Latent Class Analysis (LCA) outperforms the standard maximum likelihood LCA in predictive performance and is more robust for different number of clusters.

1 Introduction

The recent trend of globalization facilitated by world-wide communication- and Internet technologies has affected people's identity- and value formation, since common values can now be easily shared across geographical boundaries. This implies that cultural values within common regions and nations become differentiated if traditional local values are being mixed with universal values promoted by the fast paced globalization. Investigating the heterogeneities between and within nations has become a prominent topic among researchers in various disciplines such as cross-cultural psychology, sociology, and marketing sciences.

The emerging heterogeneities are often investigated by utilizing various clustering methods — automated search procedures for partitioning a data set into groups of similar data points. In practice, clustering is often based on heuristic methods (Fraley and Raftery, 2002), where the data is partitioned in order to maximize the between-cluster differences and/or minimize the within-cluster differences according to a given cost function. A popular example of this class of algorithms are centroid based clustering, such as k-means, in which data points are iteratively reallocated to clusters until no further improvement can be obtained. Another example is hierarchical agglomerative clustering, in which pairs of clusters are iteratively merged in order to optimize the chosen criterion, often being the shortest or average distance between clusters or the within-cluster

*This is an original manuscript of an article published by Taylor Francis in The Journal of Mathematical Sociology on 17 September 2020, available online: `http://www.tandfonline.com/10.1080/0022250X.2020.1811277`.

variance (Ward, 1963). Heuristic methods can be both conceptually intuitive and simple to apply, and often have very reasonable computational times. These methods, however, lack a statistical foundation, which limits the way relevant questions, such as determining an appropriate number of clusters, can be theoretically evaluated (Picard, 2007).

Probabilistic based clustering methods exists, including mixture models such as Latent Class Analysis (LCA) (McCutcheon, 1987), which has become one of the most widely used tools for conducting clustering analysis within many different research fields (Berzofsky et al., 2014). In the social sciences, LCA has been used to extract different patterns of people's behaviors, attitudes or value priorities (Szakolczai and Füstös, 1998; Magun and Rudnev, 2008, 2015; Moors and Vermunt, 2007). In sociology and cross-cultural psychology, LCA has been applied to analyze patterns of various survey responses such as the European Social Survey and the World Value Survey (Eid et al., 2003; Magun et al., 2015; Kankaraš et al., 2010; Finch and Bronk, 2011; Rudnev et al., 2014). Magun et al. (2015) applied an extended version of the traditional LCA, the so-called Factor Mixture Model (Muthén, 2008) to analyze the response patterns of 21 question items in the Portrait Value Questionnaire (PVQ21) (Schwartz et al., 2001) available from the 4th round of the European Social Survey 2008-2009 (ESS) for 29 European countries (Jowell et al., 2007), and identified five clusters based on data from approximately 55 000 respondents.

In the aforementioned works, the number of clusters is most often identified based on a model selection criterion such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC). Since a more complex model with more latent classes can always fit the data better, these criteria make a trade-off between the model fit and the complexity of the model in terms of the number of parameters, to avoid overfitting. The BIC is derived by an asympotic expansion of the posterior probability that the model is true, and assumes that the true model is among the set of models under consideration, whereas the AIC is an asymptotic estimate of the prediction error on held-out data. Which of these criteria to apply thus depends on which statistical assumptions that are made as well as on which scientific questions that are addressed. While both of these criteria are valid asymptotically under their individual assumptions, it is debatable which criterion to use in practice. Nylund et al. (2007) compares several approaches to estimating the number of clusters in LCA and similar mixture models, and conclude that the (very computationally demanding) bootstrap likelihood ratio test (BLRT) performs best, followed by BIC; however, it is clear that the model selection criterion is important and has a strong influence on the estimated number of clusters.

One of the purposes in value segmentation is to identify subgroups of individuals who share behaviors or attitudes in a manner, such that they can be characterized as accurately as possible by their latent class membership. From this point of view, it is important to estimate an appropriately large number of clusters of which members share homogeneous response patterns that are clearly distinguished from patterns indicated by other clusters. On the other hand, the *"assumption of within-segment homogeneity may be overly restrictive and result in a loss of explanatory and predictive power"* (Allenby et al., 1998). In other words, it is difficult to identify an appropriate number of clusters that are both specific and homogeneous, yet not so restrictive that their explanatory power is lost.

In an exploratory latent class analysis, the objective is often to identify a single clustering that best complies with the data, in order to provide a directly interpretable characterization of the response patterns in the data. In the classical maximum likelihood (ML) approach to LCA, the final result is the clustering which maximizes the probability of the data under the model. An issue with the ML point estimate is that it does not directly take into account the statistical uncertainty associated with the solution. If the number of clusters is low, such that each cluster has a substantial number of data associated with it, the clusters will be statistically well defined, whereas if the number of clusters is high relative to the number of data, the uncertainty may be substantial. Both Bayesian and Frequentist approaches can be used to quantify the statistical uncertainty in LCA. In the Frequentist setting, one can compute confidence sets for example using bootstrap methods, and in the Bayesian setting, one can compute a posterior distribution over clusterings using methods such as Markov chain Monte Carlo. While both have their merits, in the following we focus on the Bayesian setting. In our view, a major advantage of the Bayesian approach is that it allows us directly to compute posterior probabilities, for example to evaluate the probability that two observations belong to the same cluster or that a particular clustering is true according to the model. If needed, the Bayesian posterior can still be summarized by a single clustering that is Bayes-optimal according to a specified utility function (Rastelli and Friel, 2016); however, in terms of explanatory power, utilizing the uncertainty by averaging over the posterior distribution often yields better predictions.

In this paper we compare the classical maximum likelihood LCA (Lanza et al., 2007) with a Bayesian LCA in which the posterior uncertainty of the latent classes is taken into account. To simplify the presentation, we limit the discussion to binary data, but we note that extensions to categorical, ordered categorical, and nominal data is possible. We apply the ML and Bayesian LCA models to a human value questionnaire data set similar to the data analyzed by Magun et al. (2015). We demonstrate how prediction on held-out data can be used to compare the quality of the clusterings obtained by the LCA techniques as well as clusterings based on demographic markers. We show that the maximum likelihood and Bayesian approaches lead to similar clusterings, but that the Bayesian approach has superior predictive performance on held-out data because it incorporates uncertainty in its estimate. We further demonstrate how the Bayesian LCA can be used in a predictive approach to model order selection, in which an appropriate number of clusters is identified by optimizing the predictive performance on held-out data.

2 Data and method

Values play a central role for explaining individuals' belongings to social groups and the motivational basis of attitudes and behavior. Values have been studied by researchers in sociology, psychology and anthropology for portraying societies, organizations, and individuals.

The Schwartz dataset

Schwartz's theory of 10 basic values is one of the most widely applied value theories, and has been integrated among others in the World Value Survey (WVS) and the European

Social Survey (ESS). According to Schwartz (2012), his theory is capable of capturing characteristics of value priorities both at an individual and at a societal level, and it also accommodates primary characteristics of values previously defined by prior theorists (e.g. Allport 1961; Feather 1995; Kluckhohn 1951; Morris 1956; Rokeach 1973).

The 10 basic values (Schwartz, 2012; Smith and Schwartz, 1997) are listed in the following: *self-direction*, *stimulation*, *hedonism*, *achievement*, *power*, *security*, *conformity*, *tradition*, *benevolence* and *universalism*. In Schwartz theory, these 10 basic values are organized in a circular model indicating two aspects of value relations: conflict values vs. congruent values (see Figure 1). For example, a person with a higher stimulation value who seeks an exciting and varied life may likely undermine the tradition- or security values (conflict relation). On the other hand, a person who prioritizes the achievement value in his/her life may prioritize the power value too (congruent relation). In Figure 1, the stimulation value and the tradition- or security values are located in opposing positions, whereas the achievement- and power values are next to each other. In other words, the circular model effectively surveys the conflict vs. congruent relations among these 10 basic values (Schwartz, 2007, 2012).

Schwartz et al. (2001) further categorizes the 10 basic values into four superordinate values as follows: *openness to change*, *self-enhancement*, *conservation* and *self-transcendence* as shown in Figure 1. Among these, the values belonging to the category *openness to change* are opposed to the values belonging to the category *conservation*. In the same way, the values in the *self-enhancement* category are opposed to the values in the *self-transcendence* category.

This theory of 10 basic values has been assessed by a number of scientists (Bilsky et al., 2011; Davidov et al., 2011) who elaborate on Schwartz's theory to a dynamic model. The dynamic underpinning of the value structure describes that the values belonging to *self-enhancement* and *openness to change* directly regulate "how one expresses personal interests & characteristics" (Schwartz, 2012), whereas the values belonging to *conservation* and *self-transcendence* regulate "how one relates socially to others and affects them" (Schwartz, 2012). In addition, the values belonging to *self-enhancement* and *conservation* are anxiety-based values that prevent loss of goals (a self-protection against threats), whereas the values belonging to *openness to change* and *self-transcendence* are anxiety-free values that promote gain of goals (self-expansion and growth) (Schwartz, 2012).

Considering these aspects of Schwartz's theory, it is expected that the explorative pattern analysis of these value priorities indicated by individuals uncover heterogeneous structures of societies that are more or less invisible for the traditional cross-cultural comparative analysis. Our quantitative data analysis employs the fourth round of the European Social Study, ESS-4: 2008-2009 (Jowell et al., 2007) accessible from the ESS organization[1], in order to be able to contrast our results with existing works such as Magun et al. (2015). The dataset contains responses from approximately 55 000 respondents from 29 European countries. The questionnaire includes a simplified version of the Portrait Values Questionnaire (PVQ) developed by Schwartz et al. (2001) that consists of 21 question items portraying people expressing different goals, aspirations, or wishes that point implicitly to the importance of a value (Schwartz, 2012).

[1] http://www.europeansocialsurvey.org/

Specifically, 21 questions are classified into 10 basic values as follows: *Self-Direction* (Important to think new ideas and being creative, Important to make own decisions and be free); *Stimulation* (Important to try new and different things in life, Important to seek adventures and have an exciting life); *Hedonism* (Important to have a good time, Important to seek fun and things that give pleasure); *Security* (Important to live in secure and safe surroundings; Important that government is strong and ensures safety); *Conservation* (Important to do what is told and follow rules; Important to behave properly); *Tradition* (Important to follow traditions and customs; Important to be humble and modest, not draw attention); *Benevolence* (Important to help people and care for others well-being; Important to be loyal to friends and devote to people close); *Universalism* (Important to understand different people; Important to care for nature and environment; Important that people are treated equally and have equal opportunities); *Achievement* (Important to show abilities and be admired; Important to be successful and that people recognize achievements); *Power* (Important to be rich, have money and expensive things, Important to get respect from others).

From the original dataset accessible from ESS-4, respondents missing a response to any of these 21 question items were removed, which resulted in 51 641 respondents used in our analysis. We randomly select 80 percent of the respondents as training data with the remaining 20 percent used as hold out data for evaluating the predictive performance of the models. The answers to the 21 questions in the ESS-questionnaire are given by the following six ordered categories:

$$
\left.\begin{array}{ll}
1, & \text{Very much like me.}\\
2, & \text{Like me.}\\
3, & \text{Somewhat like me.}\\
4, & \text{A little like me.}
\end{array}\right\}\quad 1,\ \ \text{Positive response}
$$

$$
\left.\begin{array}{ll}
5, & \text{Not like me.}\\
6, & \text{Not at all like me.}
\end{array}\right\}\quad 0,\ \ \text{Negative response}
$$

Here categories 1 through 4 semantically represents positive responses while category 5 and 6 represents negative responses (see also Glückstad et al. 2016). As shown in Table 1, the asymmetry between positive and negative responses in the data is pronounced, with an average of 87 percent positive responses across all question items. As our primary objective is to highlight differences between the maximum likelihood and Bayesian LCA methods, we limit the discussion to the corresponding models for binary observations. We find it reasonable to binarize the data with the threshold for the two categories set to separate between positive and negative responses, as argued in Glückstad et al. (2016). This allows us to analyze the data using binary LCA methods, but we emphasize that similar results could be obtained by analyzing the ordered categorical data.

2.1 Latent class analysis

Within social sciences, observed data often exhibit some form of heterogeneity even though the underlying source cannot be observed directly. The goal of LCA is then to partition the population of data items into groups of similar items, based on the latent concepts that cause observed correlations within the data. For the ESS-data, the clustering problem becomes to split the N respondents into K clusters, based on the structure of their response patterns for the $Q = 21$ questions. The binarized ESS data can be considered as a binary matrix with with N rows (the respondents) and 21 columns (the question items):

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,21} \\ x_{2,1} & x_{2,2} & \dots & x_{2,21} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,21} \end{bmatrix}. \tag{1}$$

We examine and compare two LCA approaches for modeling the ESS-data:

1. The classical maximum likelihood LCA (McCutcheon, 1987) using the expectation-maximization (EM) algorithm, which has proven useful in many fields (Berzofsky et al., 2014). We utilize the SAS implementation of LCA presented in (Lanza et al., 2007).

2. A Bayesian version of LCA, particularly a finite Bayesian mixture model (BMM) with Bernoulli likelihood and Beta prior, using a Markov chain Monte Carlo procedure for inferring the posterior distribution.

In the LCA models, the data items are assumed being generated from a mixture of probability distributions, where each mixture component represents a latent cluster. In the case of binary observations, the likelihood can be expressed in terms of the latent class memberships, z_n and the response probabilities $\eta_{k,q}$ as

$$p(\boldsymbol{X}|\boldsymbol{\eta}, \boldsymbol{z}) = \prod_{n=1}^{N} \prod_{q=1}^{Q} \text{Bernoulli}(x_{n,q}; \eta_{z_n,q}), \tag{2}$$

where N is the number of respondents, Q is the number of questions, and z_n is the latent cluster assignment of respondent n. Thus, each binary response, $x_{n,q}$ is modeled by a Bernoulli distribution (biased coin flip) with parameter $\eta_{z_n,q}$ specific for the particular question and latent cluster. Equivalently, the likelihood can be expressed in terms of the probability of membership of each latent class, γ_k,

$$p(\boldsymbol{X}|\boldsymbol{\eta}, \boldsymbol{\gamma}) = \prod_{n=1}^{N} \left[\sum_{k=1}^{K} \gamma_k \prod_{q=1}^{Q} \eta_{k,q}^{x_{n,q}} (1 - \eta_{k,q})^{1-x_{n,q}} \right]. \tag{3}$$

The maximum likelihood estimate of the LCA is usually computed using an EM procedure which iteratively optimizes the expected log-likelihood function.

2.2 Bayesian LCA

The Bayesian LCA differs from the classical LCA by introducing prior distributions on the parameters. Here, we choose vague (non-informative) flexible priors, which only influence the results minimally. For the response probability parameters η, we use a separate Beta distribution for each question. The Beta distribution has two so-called hyper-parameters, which we denote β_q^+ and β_q^-, that flexibly can specify a suitable distribution on $\eta_{k,q} \in [0, 1]$. To choose the hyper-parameters, we take a hierarchical Bayesian approach, and endow them with a vague hyper-prior $p(\beta_q) \propto 1/\beta_q$. This allows us to effectively let the data define appropriate prior distributions for the response probabilities for each question and for each cluster. An advantage of the Beta priors is that the Beta distribution is conjugate to the Bernoulli likelihood, which makes it possible to analytically marginalize (integrate out) the parameters, significantly simplifying the inference procedure. For the cluster assignments z we use a standard Dirichlet-Categorical prior, which has a single hyper-parameter, the so-called concentration parameter α that governs the cluster size distribution. Again, we take a hierarchical Bayesian approach, and endow α with a vague prior, $p(\alpha) \propto 1/\alpha$. Since the Bayesian LCA allows clusters to be empty, the effective number of clusters is governed by the (maximum) number of clusters K as well as the concentration parameter α, and we note that when K is large, the prior on α will contribute to reducing the effective number of clusters.

The Bayesian LCA model can be summarized by the following generative process:

$$
\begin{aligned}
z_n &\sim \text{Dirichlet-Categorical}(\alpha) && \text{Clustering of respondents} && (4)\\
\eta_{k,q} &\sim \text{Beta}(\beta_q^+, \beta_q^-) && \text{Response probability} && (5)\\
x_{n,q} &\sim \text{Bernoulli}(\eta_{z_n q}) && \text{Response} && (6)
\end{aligned}
$$

The model parameters are inferred using Markov chain Monte Carlo, simulating samples from the posterior distribution $p(\boldsymbol{z}|\boldsymbol{X})$. The cluster labels $\boldsymbol{z}$ are inferred using Gibbs sampling, while the hyper-parameters $\alpha, \beta_1^+, ..., \beta_Q^+, \beta_1^-, ..., \beta_Q^-$ are individually inferred using a Metropolis-Hastings sampling procedure. Technical details regarding the model specification and inference procedure are described in the Appendix.

2.3 Evaluating predictive performance

A key advantage of using a generative probabilistic approach to clustering is that the model provides a principled approach to evaluating the model fit by prediction on held-out test data. Given a model fitted on data $\boldsymbol{X}$, the predictive likelihood of the 21 binary observations $\boldsymbol{x}^*$ from a new respondent is given by

$$p(\boldsymbol{x}^*|\boldsymbol{z}, \boldsymbol{X}) = \sum_{k=1}^{K} \hat{\gamma}_k \prod_{q=1}^{Q} \hat{\eta}_{k,q}^{x_q^*}(1 - \hat{\eta}_{k,q})^{1-x_q^*}, \tag{7}$$

$$\hat{\eta}_{k,q} = \frac{n_{k,q} + \beta_q^+}{m_k + \beta_q^+ + \beta_q^-}, \quad \hat{\gamma}_k = \frac{m_k + \frac{\alpha}{K}}{N}, \tag{8}$$

where $n_{k,q}$ is the number of positive responses in cluster k on feature q, and m_k is the size of cluster k. As a measure of model fit, we average the logarithm of this expression over the held-out test observations, to yield an estimate of the predictive log-likelihood. The predictive log-likelihood can be used to estimate the appropriate number of clusters, by fitting models with a varying number of clusters and comparing their predictive power.

3 Results and Analysis

Segmentations of respondents can be obtained in multiple ways. First we explore the data by partitioning the respondents according to the demographics parameters within the dataset itself. This is illustrated using respondents age, country of origin and combinations thereof. We evaluate how well groups identified by this approach captures characteristics of shared value priorities as expected (Schwartz, 2003).

Using the predictive framework, the demographics based segmentations are compared with the data-driven modelling techniques. We use the predictive log-likelihood on held-out data as the measure to compare the predictive performance of the different clustering techniques. Segmentations obtained by standard maximum likelihood and Bayesian LCA are furthermore compared both in terms of their predictive performance and according to their capability of partitioning the respondents according to their response patterns, value priorities and demographics.

3.1 Segmentations based on demographics

Based on the respondents' answers to the 21 questions, they can individually be positioned on the two value dimensions spanned by the Schwartz circle, as illustrated in Figure 2 for all respondents in the training data. The position on the horizontal axis is computed as the sum of positive responses to the 6 questions associated with the *openness to change* value group subtracted by the sum of positive responses to the 6 questions associated with the *conservation* value group. The position on the vertical axis is likewise computed as the sum of positive responses to the 5 question associated with *self-transcendence* subtracted by the sum of positive responses to the 4 questions associated with *self-enhancement*. The figure is coloured to indicate the number of respondents that share the same position in both dimensions.

The most common shared position is (0,1), corresponding to an equal number of positive responses to the *conservation* and *openness to change* value groups and one more positive answer in the *self-transcendence* value group than in *self-enhancement*. Because the *self-transcendence* value group is associated with one more question item than *self-enhancement* respondents that answers positive to all 21 questions will be positioned here.

Schwartz (2003) expects positive correlation between age and conservation values, i.e., older people tend to have stronger conservation values and self-transcendence values and vice versa (see also Tyler and Schuller 1991; Veroff et al. 1984). This effect is clearly shown in Figure 3 where the respondents are partitioned according to age-groups. The position on the value dimensions are here computed as the average response for all respondents in a given age-group. In the figure, the age-groups are further subpartitioned according to four geographical regions; *Nordic countries*, *West European countries*, *Mediterranean countries* and *Post-Communist countries* as classified by Magun et al. (2015). The geographical subpartitioning illustrates that the diversity of value-orientation is similar within all regions: The young age-groups share values associated with personal focus while older age-groups share values associated with social focus.

Although the figure shows that there exists between-region similarities on the personal to social focus diagonal (that can be explained by age), it indicates that there is between-region diversity on the protection to growth diagonal. Figure 4 further supports that this between-region diversity can be explained by nationality. Here respondents are partitioned according to nationality only. The figure indicates that respondents from post-communist countries commonly share values associated with personal protection while respondents from west and north European countries share more growth oriented values.

From Figure 3 and Figure 4 the two demographic markers (age-group and geographical region) seem fairly capable of separating the data according to the value dimensions of Schwartz's theory.

3.2 Segmentations inferred by LCA

Both the traditional and the Bayesian LCA models were fitted to the training data for the following number of clusters $K = \{5, 10, 20, ..., 70\}$. As the result of the inference for the LCA models can be influenced by initial conditions, the models were fitted 5 times with different random initial conditions for each K, resulting in five independently inferred clusterings for each K.

To asses the stability of the solutions, Figure 5 compares the sizes of the inferred clusters, both when varying K and for two independent clusterings fitted for the same K. The figure indicates that there are slight differences in the size distribution for the same K. For $K = 5$ and $K = 10$ the distribution of cluster sizes seems to be rather similar between the two models. For higher K standard LCA seems to relatively assign more respondents to the largest cluster.

3.3 Evaluating the predictive performance

The predictive log-likelihood for various number of clusters is shown in Figure 6, where predictions were made on the 20 percent held-out data using equal hyperparameters $\alpha = \beta_q^+ = \beta_q^- = 1$ for all methods. The figure shows the average over five re-runs of both standard and Bayesian LCA for the different number of clusters. For comparison, the predictive performances of clusterings obtained from K-means (using city-block distance measure) and Hierarchical clustering (using the Ward linkage function) are shown. As a baseline, the predictive performance using a random partition is included. Furthermore, the figure shows the predictive performance of using clusterings based directly on demographic variables in terms of respondents gender (2 clusters), country (29 clusters), region (4 clusters), age (86 clusters), age-group (3 clusters), and region and age-group (12 clusters, as in Figure 3).

The predictive log-likelihood of standard LCA reaches a maximum at around $K = 20$ clusters and drops when K is increased. The predictive performance of Bayesian LCA is higher than maximum likelihood LCA. It increases until around $K = 30$ clusters and remains stable from then on, demonstrating that Bayesian LCA is less prone to over-fitting.

A key advantage of the predictive approach is that it can be used for model order selection. While the Bayesian LCA does not over-fit for larger K, the predictive log-likelihood levels off which indicates that a model of order $K = 20$ is complex enough to capture most of structure within data. In order to interpret and compare inferred clusterings, a less complex clustering might be desired. Figure 6 shows that already at $K = 5$ the LCA models infers clusterings that better describe data than segmentations obtained by heuristic based methods or demographics - even for much higher K.

All the evaluated demographic-based clusterings clearly provide better than random predictions, except for *gender* which is only slightly better than random; however, as might be expected, the predictive performance obtained using only demographics is inferior to the clustering approaches, that are fitted on training data to optimally capture the statistical structure in the data.

When creating a clustering such that the training data is partitioned according to nationality, i.e 29 clusters (which is in the vicinity of the optimal number of clusters identified by the models), the predictive log-likelihood becomes significantly lower than for the two models. The same is true for the other demographics markers, clearly indicating that the models identify information beyond demographics and that there is statistical support for this in the data.

Though both K-means and Hierarchical clustering performs significantly better than using the demographic markers they perform worse than LCA, especially for low K.

The predictive performance for the Bayesian model reaches a maximum at a higher number of clusters than for standard LCA. This indicates that the Bayesian framework allows the model to reveal a more complex structure by partitioning the data into more clusters. From the maximum, the predictive performance of the Bayesian model remains constant for higher number of clusters, while it drops for the non-Bayesian LCA, allowing it to fit the data into more clusters while not over-fitting to the training data. However Figure 5 shows that a higher number of clusters results in more small clusters, while the distribution of the larger clusters seems to remain rather constant.

Bayesian LCA allows for empty clusters and hence do not guarantee that the data will be split into all K clusters. This allows the model to be less sensitive to the selected number of clusters. If K is high enough, the Bayesian model will simply not partition the data into all the available clusters and the predictive performance will remain high.

3.4 Response patterns of LCA clusterings

Table 2 shows proportions of positive responses to the PVQ21 question items indicated by members of the respective clusters for $K = 5$. The table demonstrates that the Bayesian model has extracted clusters that are highly similar to those extracted by standard LCA. For both models, the largest cluster is dominated by positive responses to all 21 questions, while the other clusters are dominated by negative responses to the questions associated with one or two subordinate values. Cluster 2 is negative towards 'openness to change' and 'self-enhancement'. Cluster 3 is negative towards 'self-enhancement' and slightly negative towards 'conservation'. Cluster 4 is slightly negative to all subordinate values except 'openness to change'. Cluster 5 is negative towards 'openness to change' and 'self-enhancement'. For $K = 5$, all clusters have positive responses to the questions associated to the subordinate value 'self-transcendence'.

Table 3 shows proportions of positive responses for clustering with $K = 20$. Here the models can recover clusters associated to negative responses for the 'self-transcendence' subordinate value. For Bayesian LCA this is seen in cluster 14 to 20 and for standard LCA this is in cluster 13 and 16 to 20. As seen from Figure 5 these clusters are fairly small, in total consisting of 3.6% and 3.8% of all respondents respectively for the two models. In the training data just 2.14% of the respondents answers positive to only one or two of the questions associated with 'self-transcendence'. This indicates that the added complexity of $K = 20$ allows the models to contain small clusters and capture such specific response patterns. The added complexity also allows the models to identify clusters that do not contain negative responses for entire subordinate value groups, but also simply for a single or a few questions across value groups.

For both $K = 5$ and $K = 20$ a relatively large proportions of sample belongs to the first cluster which do not indicate specific value priorities.

3.5 Subordinate value positions of LCA clusterings

Based on averaging the responses for the four subordinate value groups, the clusters can be positioned on the two value dimensions of the Schwartz circle: *conservation* to *openness to change* and *self-enhancement* to *self-transcendence*.

This is illustrated in Figure 7 for Bayesian LCA with $K = 5$. This figure is comparable to Figure 3 in Magun et al. (2015) where five clusters named as "Growth", "Strong Personal Focus", "Weak Personal Focus", "Strong Social Focus" and "Weak Social Focus" are plotted. Though the first, largest cluster in Figure 7 do not indicate any specific value, the figure identifies clusters for "social focus", "strong personal focus", "weak personal focus" and "growth". Similar to Magun et al. (2015) the figure shows no clustering for the "protection" value and forms the value diagonal: "social focus" to "personal focus".

The overview of the clusters positioned on the value dimensions for various K for both models are plotted in Figure 8. The clusters are ordered according to their size, with cluster 1 containing most respondents. The clusters are positioned according to their average score with a coloured area spans the standard deviation for the cluster. The figure shows that the clusters are scattered on the three focus areas: Growth, Personal focus and Social focus similar to Magun et al. (2015). For $K = 5$ the inferred clusters are very similar for the two models. For higher K the position of the cluster centroids differs, yet the distribution of clusters seem to span similar areas of the value space and form the diagonal "Personal focus" to "Social focus".

3.6 Demographics of LCA clusterings

Figure 9 depicts how the populations in the respective 29 countries are distributed across the clusters identified by the Bayesian LCA. The countries are separated into four regions "Nordic countries", "West European countries", "Mediterranean countries" and "Post-Communist countries" as classified in Magun et al. (2015). The figure shows that the countries are internally diverse, as the individual countries populations are split across the clusters and tend to be represented in all clusters. Especially for $K = 5$ it is evident that there are regional differences. Nordic and West European countries are more represented in the growth and social focus cluster (cluster 3 and 4), while the Post-Communist countries seem to be better represented in the clusters leaning towards protection and personal focus (cluster 2 and 5). The same effect is evident for $K = 20$ where Mediterranean and Post-Communist countries are less represented in the bigger growth-oriented clusters (cluster 7 and 10).

This is also captured by the inferred clusterings, as shown in Figure 10. The figure illustrates how the age-groups are distributed across the inferred clusters. The age-groups seem to be similarly distributed when considering all countries as well as the four region individually.

For $K = 5$ the percentage of respondents in the individual clusters depends on the age-group, with the same tendency across regions. For $K = 20$ the percentage of respondents in cluster 2 differs between regions though it seems to be similar for the three age groups within all regions. This indicates that the cluster represents a grouping of respondents, who share values independent of age. Cluster 2 is positioned slightly towards *self-transcendence* while being neutral on the *openness to change* to *conservation* axis. From Table 3 we identify that cluster 2 leans slightly towards *self-transcendence* mainly as a result of negative responses to the importance of being rich question.

The inferred clusters of both Bayesian and traditional LCA are represented in all value dimensions, indicating that the models are capable of capturing both the between-country and within-country diversity simultaneously.

4 Conclusion

In this paper we have addressed the problem of comparing how well different segmentations capture the underlying structure in data, by comparing models using the presented predictive framework. This highlights the key advantage of using a generative probabilistic approach, namely that the model provides a statistically salient evaluation of model fit based on evaluating the predictive log-likelihood on hold out data.

Prediction remains the intuitive and natural data-driven approach for measuring and evaluating performance. In this paper we have demonstrated how the predictive framework benefits sociological studies, by allowing comparisons of the model fit of different segmentation methods for identifying group-structure within human values survey data from the forth round of European Social Study (ESS-4). In particular we compared the predictive performance of segmentations based on demographics and Latent Class Analysis (LCA). Though demographics can characterize some of the structure within data, LCA showed a significant better predictive performance, highlighting that groups within the data are not only based on demographics and can not be identified by demographic markers alone.

Comparing the predictive performance we found that LCA performs better than both hierarchical and K-means clustering. The Bayesian version performs best, and does not seem to overfit the training data for larger number of clusters. The the extracted clusterings of Bayesian and standard LCA are however fairly similar. The inferred clusters of both Bayesian and traditional LCA are represented in all value dimensions. LCA is capable of capturing both between-region and within-region diversity simultaneously. Interpreting the clusters in terms of demographic composition, similar correlations are identified as expected from the demographics alone. However, the predictive performance of modelling significantly outperforms simply clustering according to demographics. This indicates that the data statistically support to describe the population in greater details than simply determined by demographics, and that mixture modelling to some extend is capable of quantifying this structure.

The introduction of globalized communication technologies has provided means for common value priorities to easily be promoted across geographical boundaries. Our segmentation results appropriately reflect the trend of value priorities commonly shared across regions while still observing regional and national specific characteristics, as expected when human value priorities transcends geographical and local social boundaries and divides into more complex personality types shared across borders or cultures.

Appendix

Bayesian LCA model specification

The goal of the clustering problem is to split the N respondents into K clusters, based on the structure of their response patterns for the $Q = 21$ questions. Consider the data set represented by a binary matrix as in expression (1). The Bernoulli distribution is a probability distribution of a binary random variable x, that takes on the value 1 with probability θ and 0 with probability $1 - \theta$, resulting in the following probability density function:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x} \qquad \text{for } x \in \{0,1\} \tag{9}$$

The probability that a given respondent i answers positive to question q is set to follow the Bernoulli distribution:

$$x_{iq} \sim \text{Bernoulli}(\eta_{\ell q}), \tag{10}$$

In the simplest case, this probability can be considered to be the same and independent for all respondents and only depend on the particular question q and the latent cluster $\ell = z_i$ that the respondent belongs to; parametrised by $\eta_{\ell q}$. For K clusters, we can consider a mixture of K Bernoulli distributions, such that the likelihood of all the responses $\boldsymbol{X}$ becomes:

$$p(\boldsymbol{X}|\boldsymbol{\eta}) = \prod_{i=1}^{N}\prod_{q=1}^{Q} {\eta_{z_i,q}}^{x_{i,q}}(1-\eta_{z_i,q})^{1-x_{i,q}} = \prod_{\ell=1}^{K}\prod_{q=1}^{Q} {\eta_{\ell,q}}^{N^+_{\ell,q}}(1-\eta_{\ell,q})^{N^-_{\ell,q}}, \tag{11}$$

where $N^+_{\ell,q}$ and $N^-_{\ell,q}$ respectively denotes the total sum of positive and negative responses for all respondents in cluster ℓ to question q. Conceptually, a reasonable choice for $\eta_{\ell,q}$ should be based on the ratio of positive and negative responses to the question q for the respondents in cluster ℓ.

The Beta-distribution is given by:

$$p(\mu|\beta^+,\beta^-) = \frac{1}{\mathrm{B}(\beta^+,\beta^-)}\mu^{\beta^+-1}(1-\mu)^{\beta^--1} \quad \text{for } \mu \in [0,1]\ ,\ \beta^+,\beta^- > 0, \tag{12}$$

where B() denotes the beta function, with Γ() being the gamma function:

$$\mathrm{B}(\beta^+,\beta^-) \ = \ \int_0^1 \theta^{\beta^+-1}(1-\theta)^{\beta^--1}\, d\theta \ = \ \frac{\Gamma(\beta^+)\Gamma(\beta^-)}{\Gamma(\beta^+ + \beta^-)}$$

The prior belief in the ratio of positive and negative responses is set to follow the Beta-function, which mathematically convenient acts as conjugate prior to the Bernoulli likelihood:

$$\eta_{\ell q} \sim \mathrm{Beta}(\beta_q^+, \beta_q^-), \tag{13}$$

where the believed ratio of the response ratio only depends on the particular question q.

The conjugacy of the two distributions means that the posterior distribution of the model belongs to the same family of distributions as the Beta-prior. This conjugacy allows $\boldsymbol{\eta}$ to be analytically marginalized (integrated), revealing the following joint distribution:

$$\mathrm{Bernoulli}(\boldsymbol{X}|\boldsymbol{\eta}) \cdot \mathrm{Beta}(\boldsymbol{\eta}|\boldsymbol{\beta}^+, \boldsymbol{\beta}^-) \;=\; \prod_{\ell=1}^{K} \prod_{q=1}^{Q} \frac{\mathrm{B}(N_{\ell,q}^+ + \beta_q^+, N_{\ell,q}^- + \beta_q^-)}{\mathrm{B}(\beta_q^+, \beta_q^-)}, \tag{14}$$

Let $\boldsymbol{\pi} = \{\pi_1, ..., \pi_K\}$ denote the probability distribution for any respondent to belong to the clusters, such that $p(z_i = \ell|\boldsymbol{\pi}) = \pi_\ell$. To allow for flexible cluster sizes, the clustering $\boldsymbol{z}$ of the respondents into K clusters is based on the Dirichlet distribution:

$$p(\boldsymbol{\pi}|\boldsymbol{c}) = \frac{1}{\mathrm{B}(\boldsymbol{c})} \prod_{k=1}^{K} \pi_k^{c_k - 1}, \quad \text{where} \quad \mathrm{B}(\boldsymbol{c}) = \frac{\Gamma(\sum_{k=1}^{K} c_k)}{\prod_{k=1}^{K} \Gamma(c_k)} \tag{15}$$

With no prior information to pick one cluster above another, a symmetric distribution is preferred. With equal concentration parameters: $\frac{\alpha}{K} = c_1 = ... = c_K$, the following joint prior over $\boldsymbol{z}$ and $\boldsymbol{\pi}$ is obtained:

$$p(\boldsymbol{\pi}, \boldsymbol{z}|\boldsymbol{c}) \;=\; p(\boldsymbol{\pi}|\boldsymbol{c}) \prod_{i=1}^{N} p(z_i|\boldsymbol{\pi}) \;=\; \frac{1}{\mathrm{B}(\boldsymbol{c})} \prod_{k=1}^{K} \pi_k^{m_k + c_k - 1}, \tag{16}$$

where m_k is the number of respondents in cluster k. Marginalizing over $\boldsymbol{\pi}$ reveals the following effective prior over $\boldsymbol{z}$

$$p(\boldsymbol{z}|\alpha) \;=\; \int p(\boldsymbol{\pi}, \boldsymbol{z}|\boldsymbol{c})\, d\pi \;=\; \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \prod_{k=1}^{K} \frac{\Gamma(\frac{\alpha}{K} + m_k)}{\Gamma(\frac{\alpha}{K})}, \tag{17}$$

being the Pólya distribution depending on the single parameter α.

Finally, the joint posterior distribution of the model is obtained when joining (14) and (17):

$$\begin{aligned} p(\boldsymbol{X}, \boldsymbol{z}|\alpha, \boldsymbol{\beta}^+, \boldsymbol{\beta}^-) \quad &= \\ &\frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \prod_{k=1}^{K} \frac{\Gamma(\frac{\alpha}{K} + m_k)}{\Gamma(\frac{\alpha}{K})} \prod_{\ell=1}^{K} \prod_{q=1}^{Q} \frac{\mathrm{B}(N_{\ell,q}^+ + \beta_q^+, N_{\ell,q}^- + \beta_q^-)}{\mathrm{B}(\beta_q^+, \beta_q^-)}. \end{aligned} \tag{18}$$

Inference in the Bayesian LCA model

The model parameters are inferred by a sequence of Markov Chain Monte Carlo methods. The clustering is inferred using a procedure of both full and restricted Gibbs sampling. In the full Gibbs sampling, all respondents are iteratively proposed reassigned to the K

clusters based on the posterior distribution of the cluster-assignment for the particular respondent, obtained by Bayes' theorem for equation (18):

$$p(z_i = \ell | \boldsymbol{X}, \boldsymbol{z}^{\setminus i}, \alpha, \boldsymbol{\beta}^+, \boldsymbol{\beta}^-) = \frac{p(\boldsymbol{X}, \boldsymbol{z}^{\setminus i}, z_i = \ell | \alpha, \boldsymbol{\beta}^+, \boldsymbol{\beta}^-)}{\sum_{k=1}^{K} p(\boldsymbol{X}, \boldsymbol{z}^{\setminus i}, z_i = k | \alpha, \boldsymbol{\beta}^+, \boldsymbol{\beta}^-)} \tag{19}$$

In the restricted Gibbs sampling, two clusters are randomly selected and three Gibbs sweep are performed, restricted to re-partitioning the nodes within the selected clusters.

The model contains a number of hyper-parameters:

$$\alpha \ , \ \beta_1^+, ..., \beta_Q^+ \ , \ \beta_1^-, ..., \beta_Q^-$$

They are all sampled independently using a Metropolis-Hastings procedure. Here, proposals for each of the parameters are drawn from a Gaussian distribution centered at the current value of the parameter and with variance 1. The proposals are accepted or rejected according to the Metropolis-Hastings accepting criterion, being the ratio of how likely the model is when using the proposed parameter value compared to the current value.

For all experiments, our sampling strategy consists of 1000 sweeps of the following sampling procedures. First a complete Gibbs sweep over all respondents is performed followed by three proposals of the restricted Gibbs sampling and 10 Metropolis-Hastings proposals for each of the hyper-parameters.

References

G. M. Allenby, N. Arora, and J. L. Ginter. On the heterogeneity of demand. *Journal of Marketing Research*, 35:384–389, 1998.

G. W. Allport. Pattern and growth in personality. 1961.

M. E. Berzofsky, P. P. Biemer, and W. D. Kalsbeek. Local dependence in latent class analysis of rare and sensitive events. *Sociological Methods & Research*, 43(1):137–170, 2014.

W. Bilsky, M. Janik, and S. H. Schwartz. The structural organization of human values: Evidence from three rounds of the european social survey (ess). *Journal of Cross-Cultural Psychology*, 42(5):759–776, 2011.

E. Davidov, P. Schmidt, and J. Billiet. *Cross-cultural analysis: Methods and applications*. Routledge, 2011.

M. Eid, R. Langeheine, and E. Diener. Comparing typological structures across cultures by multigroup latent class analysis a primer. *Journal of Cross-Cultural Psychology*, 34 (2):195–210, 2003.

N. T. Feather. Values, valences, and choice: The influences of values on the perceived attractiveness and choice of alternatives. *Journal of personality and social psychology*, 68(6):1135–1151, 1995.

W. H. Finch and K. C. Bronk. Conducting confirmatory latent class analysis using m plus. *Structural Equation Modeling*, 18(1):132–151, 2011.

C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.

F. K. Glückstad, M. N. Schmidt, and M. Mørup. Examination of heterogeneous societies: Identifying subpopulations by contrasting cultures. *Journal of Cross-Cultural Psychology*, 2016.

R. Jowell, C. Roberts, R. Fitzgerald, and G. Eva. *Measuring attitudes cross-nationally: Lessons from the European Social Survey.* Sage, 2007.

M. Kankaraš, G. Moors, and J. K. Vermunt. Testing for measurement invariance with latent class analysis. *Cross-cultural analysis: Methods and applications*, pages 359–384, 2010.

C. Kluckhohn. The study of culture. in d. lerner & h.d. lasswell (eds.). *The policy sciences*, pages 86–101, 1951.

S. T. Lanza, L. M. Collins, D. R. Lemmon, and J. L. Schafer. Proc lca: A sas procedure for latent class analysis. *Structural Equation Modeling*, 14(4):671–694, 2007.

V. Magun and M. Rudnev. Basic human values of russians: similarities and dissimilarities with the other european countries. *The Russian Public Opinion Herald*, (93):33–58, 2008.

V. Magun and M. Rudnev. Basic human values of the russians. *Culture Matters in Russia—and Everywhere: Backdrop for the Russia-Ukraine Conflict*, pages 431–450, 2015.

V. Magun, M. Rudnev, and P. Schmidt. Within-and between-country value diversity in europe: A typological approach. *European Sociological Review*, pages 1–14, 2015.

A. McCutcheon. Latent class analysis. sage university paper series on quantitative applications in the social sciences, no. 07-064. 1987.

G. Moors and J. Vermunt. Heterogeneity in post-materialist value priorities. evidence from a latent class discrete choice approach. *European Sociological Review*, 23(5):631–648, 2007.

C. Morris. Varieties of human value. 1956.

B. Muthén. Latent variable hybrids: Overview of old and new models. *Advances in latent variable mixture models*, 1:1–25, 2008.

K. L. Nylund, T. Asparouhov, and B. O. Muthén. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural equation modeling*, 14(4):535–569, 2007.

F. Picard. An introduction to mixture models. *Statistics for Systems Biology, Research Report*, (7), 2007.

R. Rastelli and N. Friel. Optimal bayesian estimators for latent variable cluster models. *arXiv preprint arXiv:1607.02325*, 2016.

M. Rokeach. *The nature of human values*. Free press New York, 1973.

M. Rudnev, V. Magun, and P. Schmidt. The stability of the value typology of europeans: Testing invariance with confirmatory latent class analysis. *Higher School of Economics Research Paper No. WP BRP*, 2014.

S. H. Schwartz. A proposal for measuring value orientations across nations. *Questionnaire Package of the European Social Survey*, pages 259–290, 2003.

S. H. Schwartz. Value orientations: Measurement, antecedents and consequences across nations. *Measuring attitudes cross-nationally: Lessons from the European Social Survey*, pages 169–203, 2007.

S. H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1), 2012.

S. H. Schwartz, G. Melech, A. Lehmann, S. Burgess, M. Harris, and V. Owens. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, 32(5):519–542, 2001.

P. B. Smith and S. H. Schwartz. Values. *Handbook of cross-cultural psychology, 2nd ed., vol.*, 3:77–118, 1997.

A. Szakolczai and L. Füstös. Value systems in axial moments a comparative analysis of 24 european countries. *European Sociological Review*, 14(3):211–229, 1998.

T. R. Tyler and R. A. Schuller. Aging and attitude change. *Journal of personality and social psychology*, 61(5):689–697, 1991.

J. Veroff, D. Reuman, and S. Feld. Motives in american men and women across the adult life span. *Developmental psychology*, 20(6):1142–1158, 1984.

J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

List of Figures

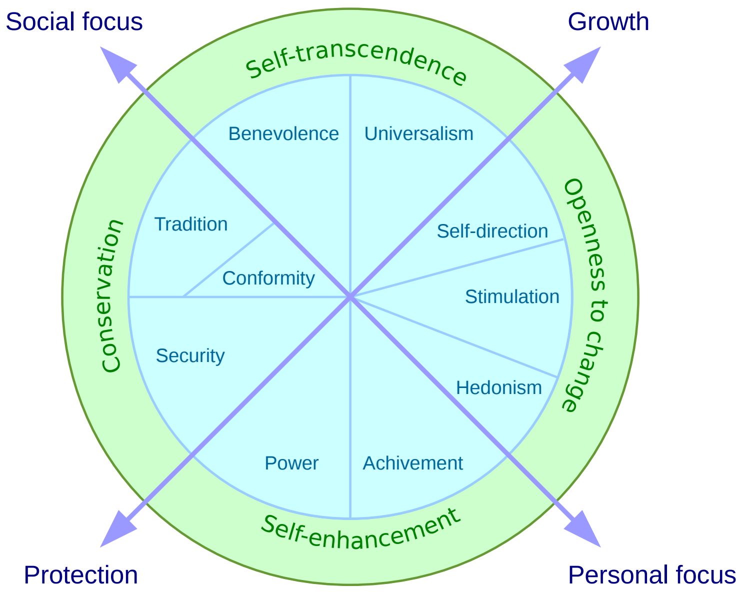

Figure 1: Circular model illustrating the relations between the 10 basic values of Schwartz' theory.

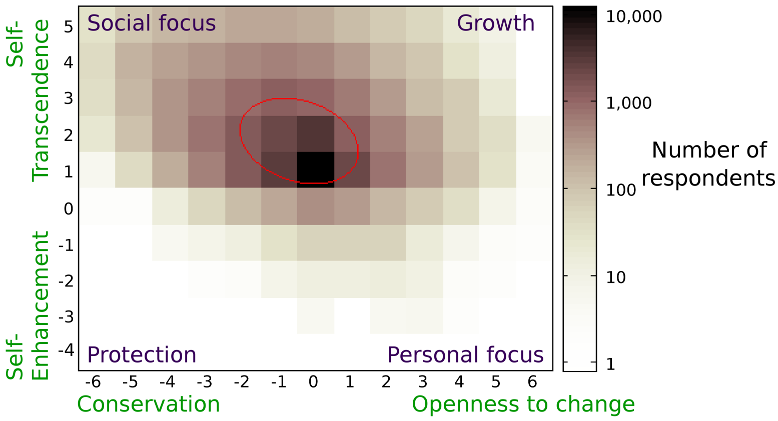

Figure 2: The entire training data projected onto the two value-dimensions of the Schwartz circle in Figure 1. Color intensity indicates number of respondents. The ellipse indicates one standard deviation from the mean of all respondents.

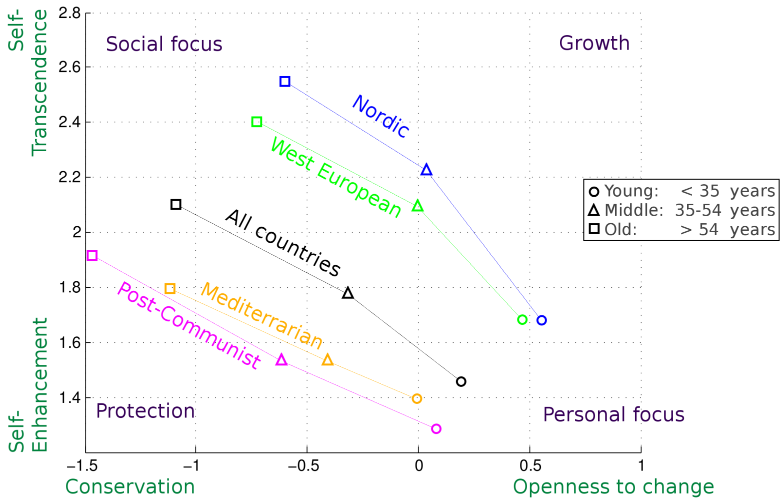

Figure 3: Clustered according to age groups (young, middle and old), the training data is projected on the value axis of the Schwartz circle shown in figure 1. Respondents in the data without an associated age are ignored.

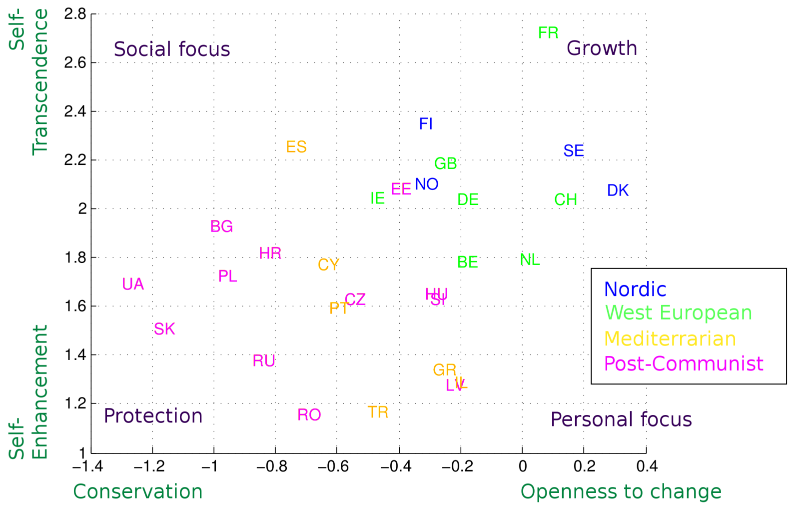

Figure 4: Clustered according to country, the training data is projected on the value axis of the Schwartz circle shown in Figure 1. Each country is positioned according to the average value for all respondents in the training data. Country names are abbreviated as in Figure 9.

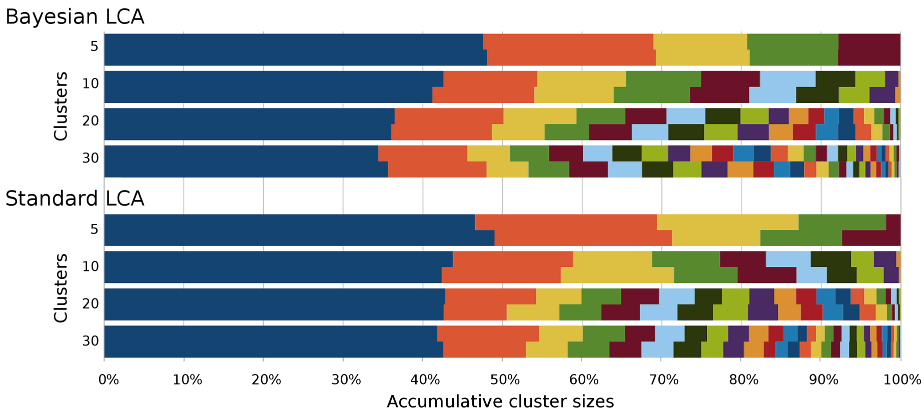

Figure 5: Stacked cluster sizes of clusterings inferred by multiple runs of the models on the training data. The figure shows results for two independent runs performed for $K = \{5, 10, 20, 30\}$ number of clusters respectively for the models.

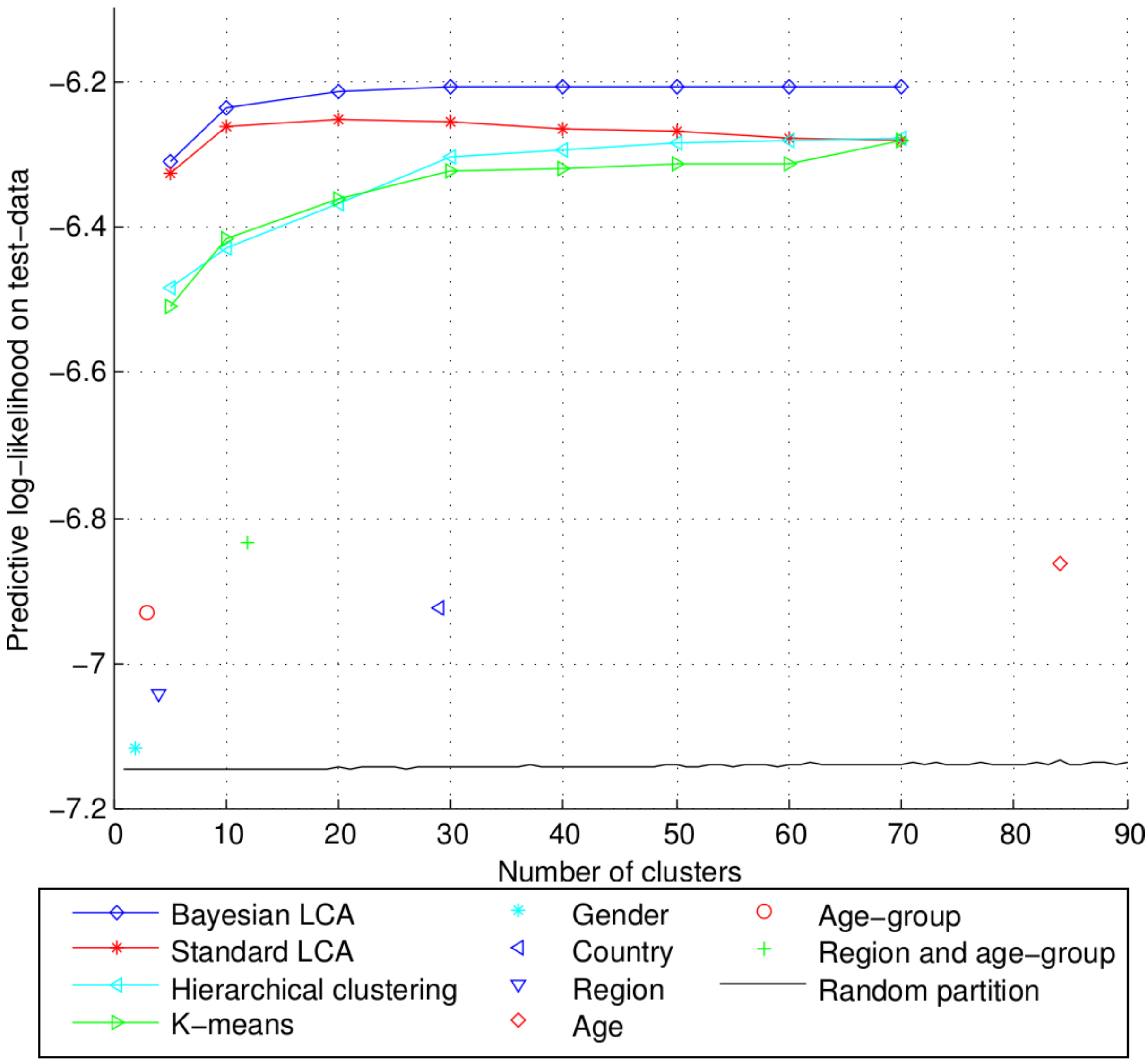

Figure 6: The average predictive log-likelihood versus the number of clusters. For the LCA models, the results are averaged over 5 random restarts of the inference procedures.

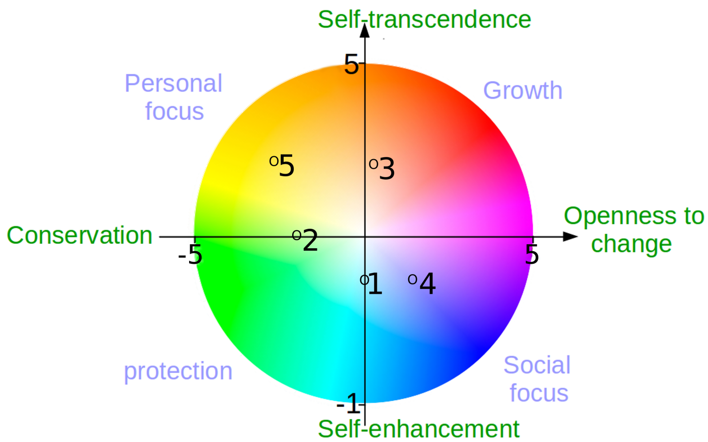

Figure 7: The clusters inferred by Bayesian LCA for $K = 5$ clusters, projected onto the two dimensions spanned by the value axis of the Schwartz circle.

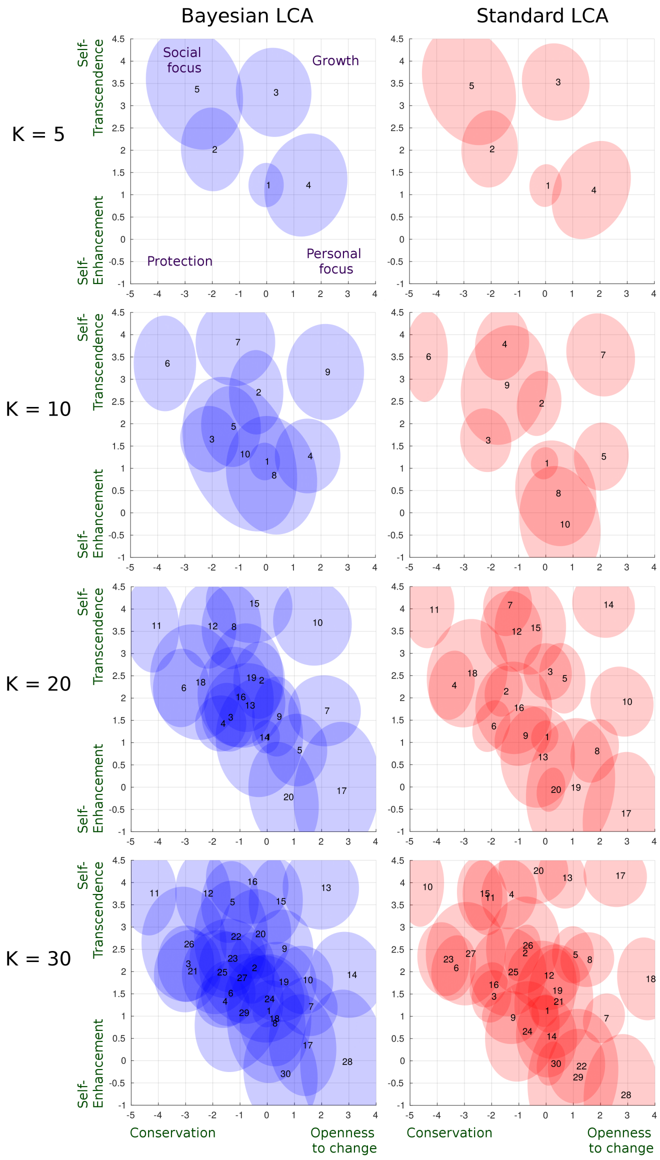

Figure 8: The extracted clusters of the two models for various K, projected on the two value axis of the Schwartz circle shown in figure 1. The clusters are positioned according to their average score and numbered according to their size. The coloured area around a number spans the standard deviation for the cluster.

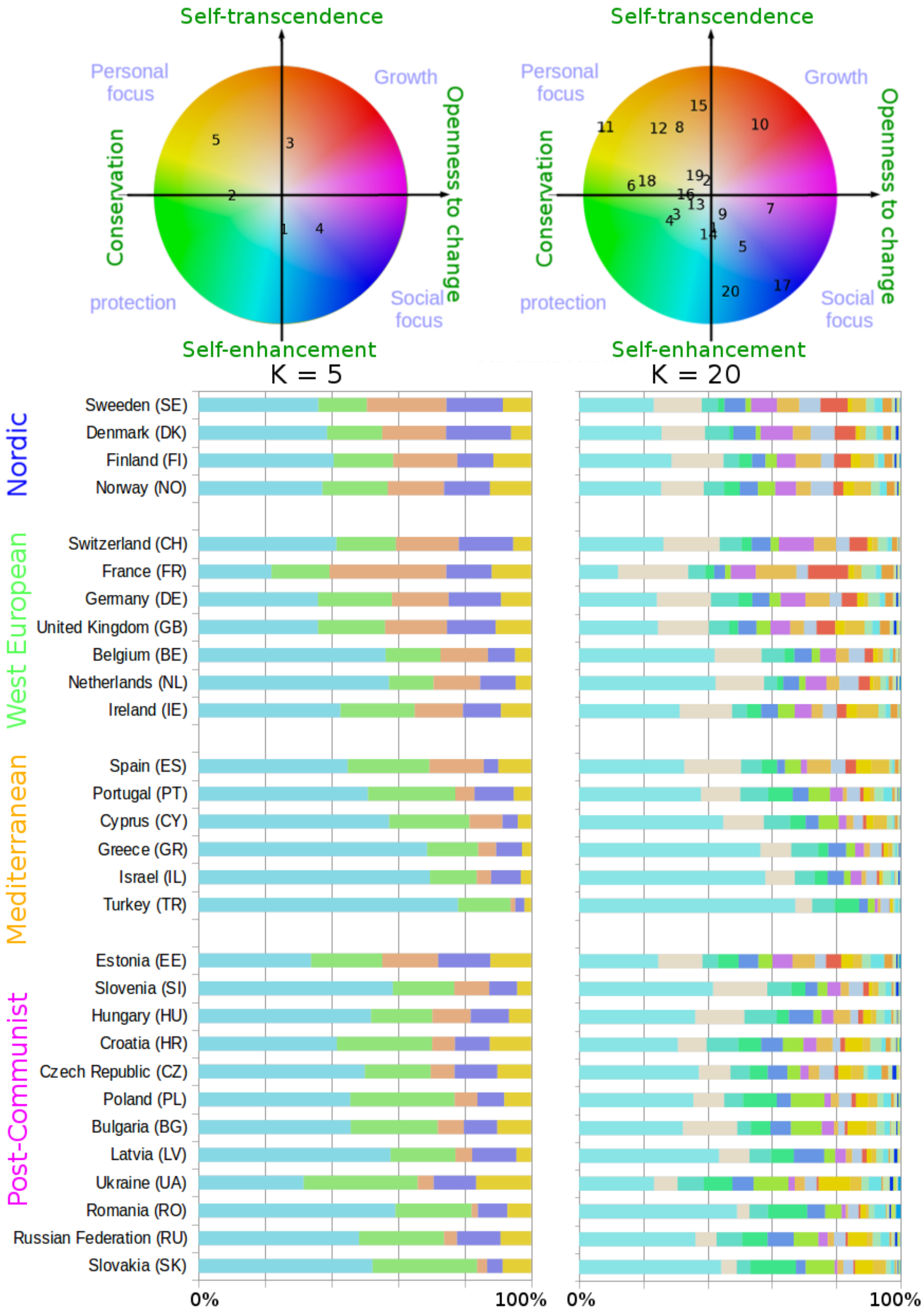

Figure 9: Distribution of the residents of the 29 countries within the inferred clusters. The figure compares Bayesian LCA with $K = 5$ and $K = 20$ and illustrates the clusters projected onto the two value dimensions of the Schwartz circle.

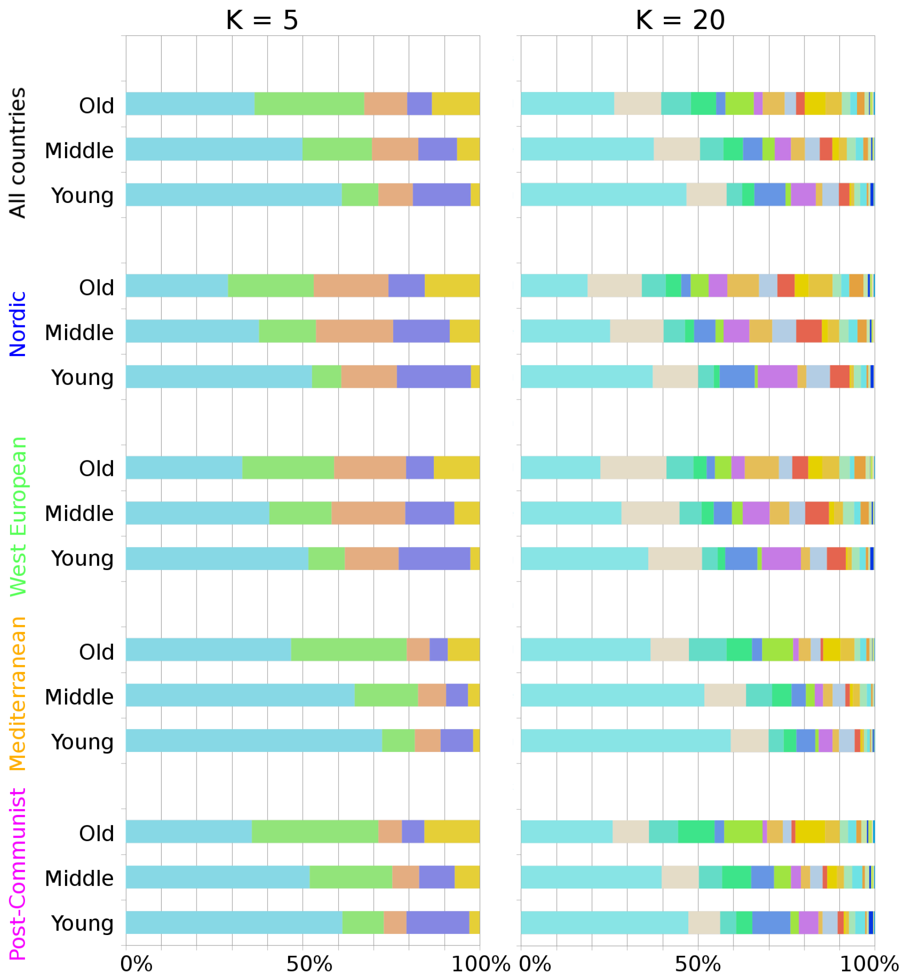

Figure 10: Distribution of age-groups within the inferred clusters. The figure compares Bayesian LCA with $K = 5$ and $K = 20$ and shows the distribution for all countries as well as the four regions individually. The figure is based on the training data, split according to age (young: younger than 35 years, middle: 35 to 54 years old, old: older than 54 years). Respondents with no associated age information in the data are ignored.

List of Tables

Value group	Question	Mean of six-scale	Percent positive
Openness to change	Creative	2.59	90.54
	New things	2.96	84.12
	Good time	2.95	84.35
	Own decisions	2.21	95.79
	Adventures	3.88	**59.25**
	Fun	3.09	81.38
Conservatism	Secure	2.27	93.81
	Follow rules	2.99	82.57
	Modest	2.72	89.19
	Safety	2.27	94.13
	Behavior	2.56	91.92
	Traditions	2.60	89.62
Self-transcendence	Equality	2.09	96.74
	Understand	2.38	95.86
	Help people	2.20	**97.73**
	Friends	**1.97**	98.48
	Nature	2.13	97.42
Self-enhancement	Rich	**3.89**	59.57
	Abilities	3.02	82.37
	Success	3.03	82.64
	Respect	3.06	81.47
Mean of all responses		2.71	87.09

Table 1: The 21 questions of the ESS dataset, split into four value groups. The table lists the mean when considering the six level answer categories as scalable continuous data and lists the percentage of positive answers for each question item.

Question \ Cluster	Bayesian LCA 1	Bayesian LCA 2	Bayesian LCA 3	Bayesian LCA 4	Bayesian LCA 5	Standard LCA 1	Standard LCA 2	Standard LCA 3	Standard LCA 4	Standard LCA 5
Openness to change										
Creative	98%	84%	94%	92%	56%	98%	85%	93%	92%	53%
New things	98%	67%	90%	88%	30%	98%	66%	91%	88%	28%
Good time	99%	66%	88%	90%	35%	98%	67%	88%	90%	31%
Own decisions	100%	95%	96%	94%	77%	100%	95%	96%	94%	75%
Adventures	85%	13%	54%	77%	7%	87%	7%	57%	79%	6%
Fun	98%	55%	90%	89%	30%	98%	55%	92%	89%	25%
Conservation										
Secure	99%	99%	83%	77%	88%	99%	99%	82%	74%	88%
Follow rules	93%	92%	62%	48%	72%	93%	92%	59%	45%	73%
Modest	94%	95%	91%	57%	88%	94%	95%	92%	53%	87%
Safety	99%	99%	87%	77%	85%	99%	99%	85%	76%	84%
Behavior	99%	98%	85%	63%	85%	99%	99%	84%	57%	84%
Traditions	96%	96%	79%	65%	82%	96%	96%	79%	61%	82%
Self-transcendence										
Equality	99%	98%	98%	88%	90%	99%	98%	97%	87%	89%
Understand	99%	97%	99%	83%	86%	99%	97%	99%	82%	86%
Help people	100%	99%	100%	90%	89%	100%	99%	100%	89%	88%
Friends	100%	100%	100%	94%	92%	100%	100%	100%	93%	92%
Nature	100%	99%	99%	88%	91%	100%	99%	99%	86%	90%
Self-enhancement										
Rich	83%	44%	14%	69%	14%	85%	40%	10%	70%	13%
Abilities	98%	82%	50%	89%	27%	99%	83%	43%	90%	23%
Success	99%	81%	52%	88%	23%	99%	82%	44%	90%	18%
Respect	95%	84%	49%	75%	47%	95%	85%	45%	76%	47%

Table 2: Proportions of positive responses to th PVQ21 question items, when modelling with $K = 5$. The clusters are ordered according to size.

Bayesian LCA

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Openness to change	Creative	99%	93%	84%	99%	97%	72%	98%	81%	94%	97%	34%	96%	86%	83%	81%	42%	69%	23%	32%	36%
	New things	99%	97%	57%	93%	95%	46%	96%	64%	93%	94%	12%	75%	50%	76%	43%	34%	82%	13%	42%	27%
	Good time	99%	92%	94%	66%	96%	33%	95%	100%	93%	89%	17%	11%	74%	61%	47%	72%	99%	11%	100%	27%
	Own decisions	100%	97%	97%	97%	97%	93%	99%	92%	99%	95%	74%	94%	93%	79%	88%	80%	93%	52%	93%	13%
	Adventures	89%	57%	15%	21%	89%	5%	87%	7%	97%	72%	0%	15%	6%	47%	11%	59%	88%	14%	22%	23%
	Fun	98%	94%	84%	35%	97%	27%	93%	73%	98%	89%	15%	46%	57%	55%	45%	86%	96%	16%	87%	14%
Conservation	Secure	100%	98%	99%	100%	96%	100%	67%	97%	82%	62%	98%	92%	93%	78%	71%	76%	56%	73%	93%	15%
	Follow rules	96%	80%	92%	95%	59%	99%	37%	76%	92%	38%	91%	83%	44%	76%	42%	58%	41%	66%	67%	9%
	Modest	96%	96%	91%	93%	57%	97%	60%	95%	84%	81%	97%	96%	75%	70%	80%	81%	44%	70%	80%	18%
	Safety	100%	98%	97%	100%	94%	99%	77%	95%	89%	71%	97%	95%	82%	72%	62%	92%	46%	63%	97%	17%
	Behavior	100%	96%	98%	98%	79%	99%	55%	97%	99%	56%	98%	97%	78%	74%	60%	85%	42%	55%	73%	12%
	Traditions	97%	91%	97%	96%	78%	97%	61%	88%	91%	61%	96%	92%	76%	61%	66%	98%	43%	66%	43%	9%
Self trans-cendence	Equality	100%	97%	98%	99%	92%	100%	97%	97%	93%	97%	95%	98%	96%	82%	95%	72%	56%	75%	90%	19%
	Understand	100%	99%	95%	99%	81%	96%	99%	97%	98%	97%	94%	98%	89%	77%	90%	67%	44%	56%	58%	17%
	Help people	100%	100%	99%	100%	94%	100%	100%	100%	97%	99%	94%	99%	97%	69%	97%	94%	62%	61%	53%	
	Friends	100%	100%	99%	99%	97%	100%	100%	99%	98%	98%	98%	99%	100%	76%	99%	98%	75%	62%	96%	14%
	Nature	100%	100%	98%	100%	94%	99%	95%	99%	96%	98%	97%	99%	90%	79%	93%	94%	44%	66%	82%	8%
Self enhan-cement	Rich	89%	28%	57%	70%	91%	34%	56%	16%	60%	13%	14%	10%	38%	74%	6%	22%	78%	5%	44%	28%
	Abilities	100%	78%	88%	96%	99%	82%	93%	30%	77%	42%	23%	44%	98%	71%	15%	40%	72%	26%	15%	26%
	Success	100%	81%	91%	100%	99%	75%	97%	29%	88%	33%	21%	26%	85%	65%	16%	86%	65%	14%	17%	24%
	Respect	97%	70%	96%	88%	86%	82%	74%	57%	97%	31%	57%	52%	67%	61%	24%	75%	74%	40%	56%	15%

Standard LCA

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Openness to change	Creative	99%	81%	95%	70%	92%	98%	94%	97%	91%	99%	39%	60%	84%	95%	98%	23%	76%	14%	59%	6%
	New things	99%	70%	100%	41%	97%	94%	93%	97%	43%	94%	9%	24%	75%	95%	54%	32%	89%	8%	51%	9%
	Good time	99%	82%	93%	28%	91%	65%	63%	98%	95%	93%	14%	99%	66%	91%	18%	63%	100%	10%	60%	0%
	Own decisions	100%	96%	97%	90%	98%	97%	92%	99%	98%	97%	73%	92%	76%	96%	89%	74%	99%	54%	42%	3%
	Adventures	87%	16%	53%	2%	97%	19%	16%	100%	8%	81%	0%	2%	54%	74%	16%	64%	95%	8%	44%	6%
	Fun	98%	98%	93%	8%	99%	11%	65%	100%	81%	88%	15%	77%	62%	95%	24%	82%	100%	15%	35%	3%
Conservation	Secure	99%	100%	100%	99%	79%	100%	94%	89%	97%	49%	98%	93%	86%	58%	69%	67%	66%	74%	36%	0%
	Follow rules	95%	100%	55%	95%	83%	99%	84%	49%	66%	23%	92%	55%	78%	29%	44%	45%	35%	67%	32%	3%
	Modest	95%	99%	97%	97%	89%	92%	98%	44%	70%	55%	98%	91%	69%	79%	69%	75%	35%	68%	39%	3%
	Safety	99%	100%	99%	99%	86%	100%	97%	92%	89%	60%	98%	85%	77%	64%	53%	91%	59%	63%	40%	0%
	Behavior	100%	100%	91%	98%	97%	100%	98%	65%	88%	36%	98%	86%	74%	50%	60%	82%	47%	54%	32%	0%
	Traditions	98%	100%	83%	96%	81%	95%	95%	71%	90%	50%	96%	72%	64%	56%	63%	98%	41%	74%	21%	3%
Self trans-cendence	Equality	99%	98%	98%	100%	94%	99%	99%	92%	96%	95%	96%	93%	84%	98%	90%	55%	51%	79%	39%	6%
	Understand	100%	100%	100%	95%	99%	100%	100%	84%	82%	99%	96%	90%	74%	95%	90%	50%	26%	47%	43%	3%
	Help people	100%	99%	100%	99%	98%	100%	100%	96%	100%	100%	95%	95%	64%	99%	95%	93%	46%	54%	29%	6%
	Friends	100%	100%	100%	99%	100%	99%	99%	97%	100%	100%	98%	100%	71%	97%	98%	96%	89%	63%	26%	0%
	Nature	100%	100%	100%	98%	97%	100%	100%	92%	96%	93%	98%	95%	75%	97%	94%	94%	43%	66%	21%	0%
Self enhan-cement	Rich	89%	20%	24%	28%	29%	75%	9%	94%	74%	42%	12%	15%	79%	12%	13%	20%	86%	2%	42%	12%
	Abilities	99%	80%	90%	86%	55%	98%	21%	100%	94%	92%	7%	30%	[illegible]	31%	36%	29%	76%	15%	52%	3%
	Success	100%	85%	91%	68%	65%	100%	20%	100%	96%	97%	13%	28%	72%	14%	28%	84%	71%	3%	37%	6%
	Respect	98%	98%	34%	80%	95%	89%	39%	86%	95%	64%	53%	51%	65%	20%	33%	77%	82%	32%	28%	0%

Table 3: Proportions of positive responses to th PVQ21 question items, when modelling for $K = 20$. The clusters are ordered according to size.