Appendix

"What Do People Want?"

Daniel J. Benjamin

Kristen B. Cooper

Ori Heffetz

Miles Kimball

Tushar Kundu

May 15, 2025

Table of Contents

A1 Computational Text Analysis Methodology

To create interpretable, data-driven aspect groupings, we developed a multi-stage computational analysis pipeline that leverages natural language processing (NLP) and machine learning techniques. This appendix details our methodology.

A1.1 Generating Standardized Aspect Descriptions

First, we created standardized descriptions of each aspect to establish a common semantic basis for comparison. Using the GPT-4 language model (OpenAI), we generated concise two-sentence descriptions for each of our 126 aspects. This step was necessary because the aspect names alone (e.g., "Your children's health," "Your financial security") lacked sufficient textual context for robust semantic analysis.

We instructed the model to "describe and explain the aspect in two sentences, capturing what it means to most people in the current time," and specified that descriptions should be written to ensure anyone reading them "understands the essence of the aspect, and what it means to most people in the world today." This approach standardized the semantic representation of aspects while preserving their core meaning.

A1.2 Creating Semantic Embeddings

Next, we transformed these text descriptions into high-dimensional numerical representations (embeddings) using a state-of-the-art sentence transformer model (all-mpnet-base-v2). This model converts text into 768-dimensional vectors that capture semantic relationships between aspects, with similar concepts positioned closer together in the vector space.

The embedding process used the SentenceTransformer framework and involved:

1. Tokenizing each description into word pieces
2. Passing these tokens through the pre-trained neural network
3. Generating a dense vector representation for each aspect

These embeddings effectively distilled complex natural language descriptions into mathematical representations that preserve semantic relationships between aspects.

A1.3 Similarity Analysis and Dimensionality Reduction

We computed cosine similarity between all aspect pairs to create a similarity matrix, which quantifies the semantic relatedness between aspects. To visualize these relationships, we applied Uniform Manifold Approximation and Projection (UMAP) to reduce the 768-dimensional embeddings to a two-dimensional representation while preserving local and global structure. UMAP parameters were tuned to capture both local relationships (n_neighbors=5) and global structure (min_dist=0.3). This dimensionality reduction enabled the visualization of semantic relationships between aspects in a two-dimensional space.

A1.4 Cluster Analysis

To identify coherent aspect groups, we applied Spectral Clustering to the similarity matrix. This algorithm is particularly effective for detecting clusters in data represented by similarity graphs. We explored different clustering configurations to determine the optimal number of clusters (n=15) that balances granularity with interpretability.

The clustering process:

1. Normalized the similarity matrix to ensure proper scaling
2. Applied Spectral Clustering using the similarity matrix as input
3. Assigned each aspect to one of 15 clusters

A1.5 Cluster Interpretation and Naming

To interpret each cluster, we employed two complementary approaches:

1. **TF-IDF Analysis**: We identified the most distinctive terms within each cluster using Term Frequency-Inverse Document Frequency (TF-IDF) analysis. This helped extract key words that characterize each aspect group while downweighting common terms that appear across many aspects.

2. **Topic Modeling**: We applied BERTopic, a topic modeling technique that leverages transformers and class-based TF-IDF, to identify coherent themes within clusters. We used a custom set of stop words to filter out generic terms and extract meaningful topics.

In practice, BERTopic failed to generate topic names in most cases. Therefore, the resulting clusters were named using a combination of the top terms identified by TF-IDF analysis and subjective judgement to label the underlying conceptual grouping.

A2 Marginal Utility Estimates by Group
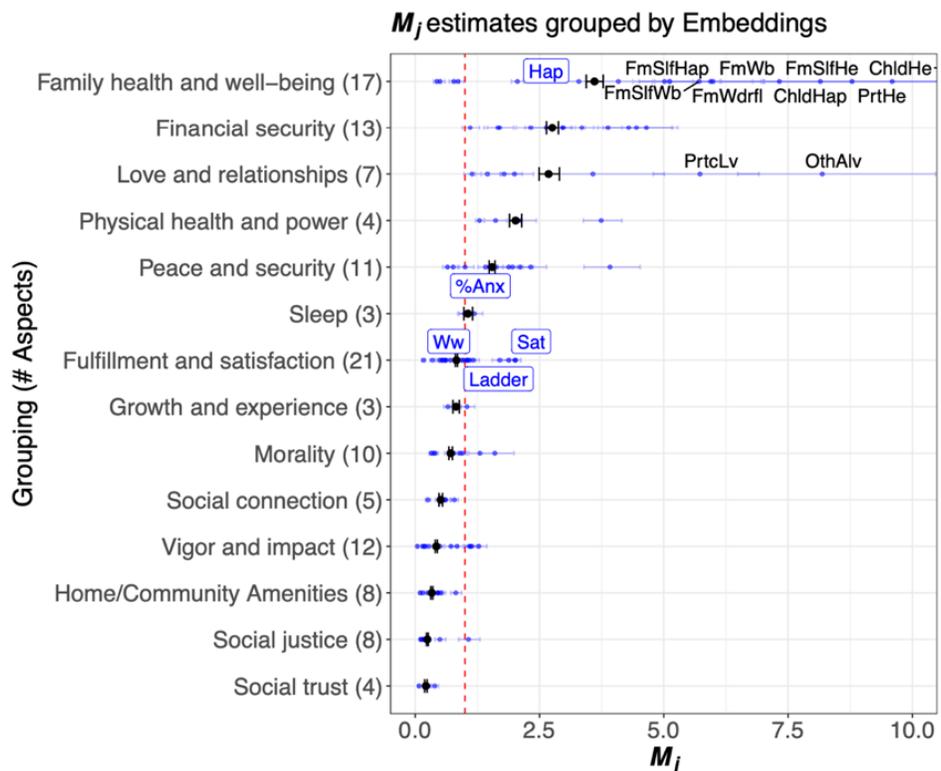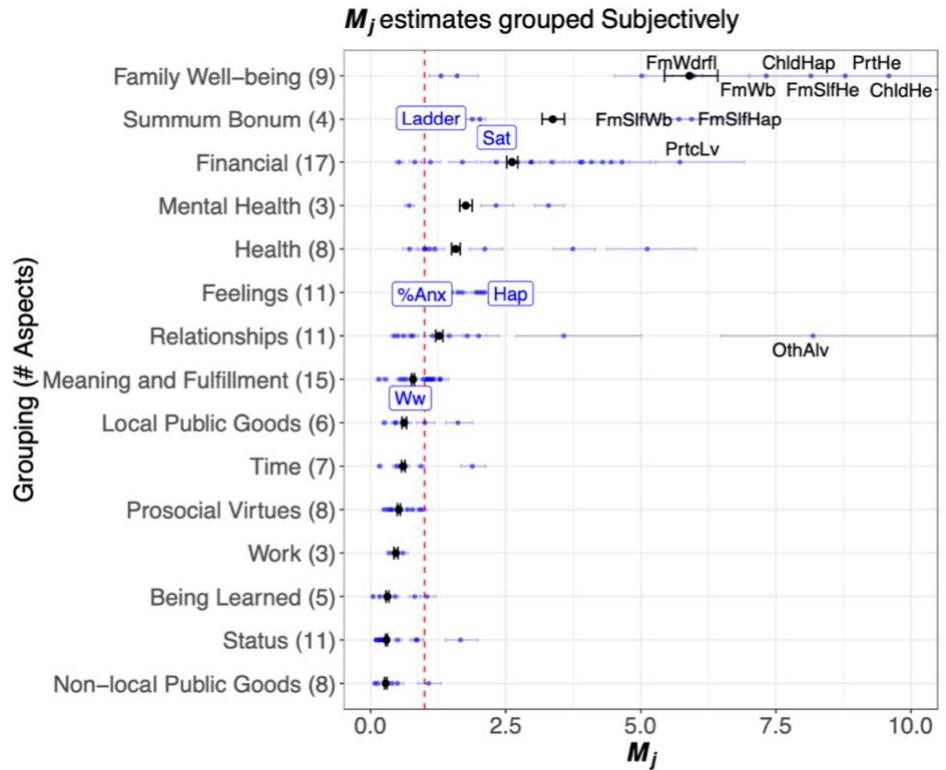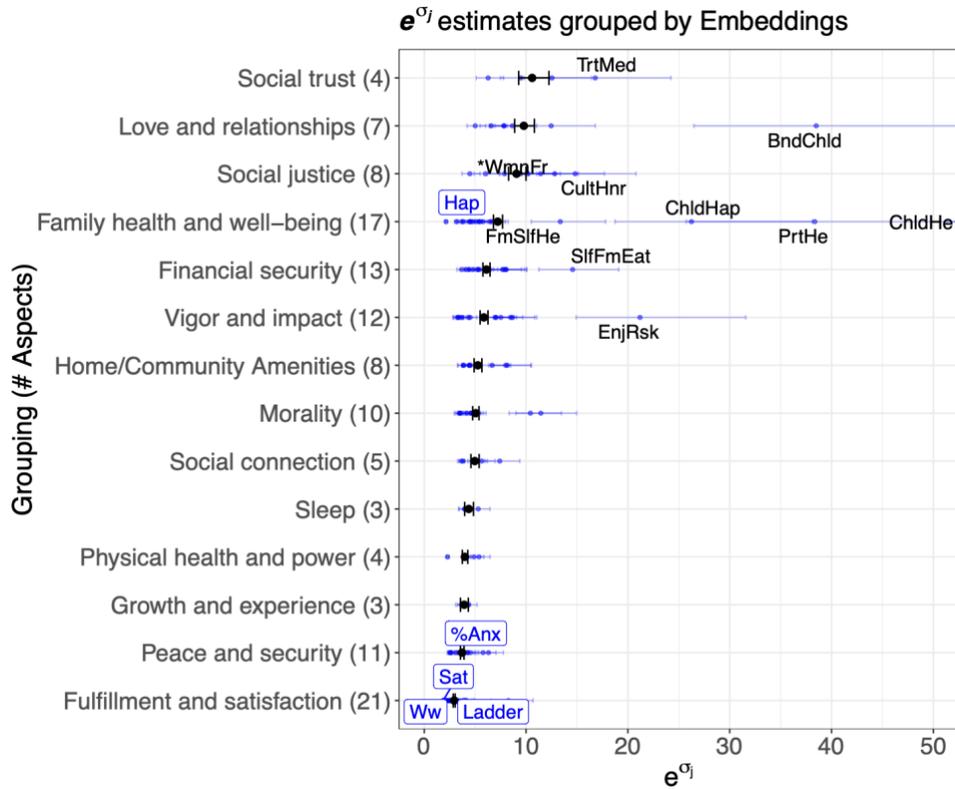
Figure A1: M_j By Group

M_j estimates grouped Subjectively

M_j estimates grouped by Embeddings

Figure A2: e^{σ_j} by Group

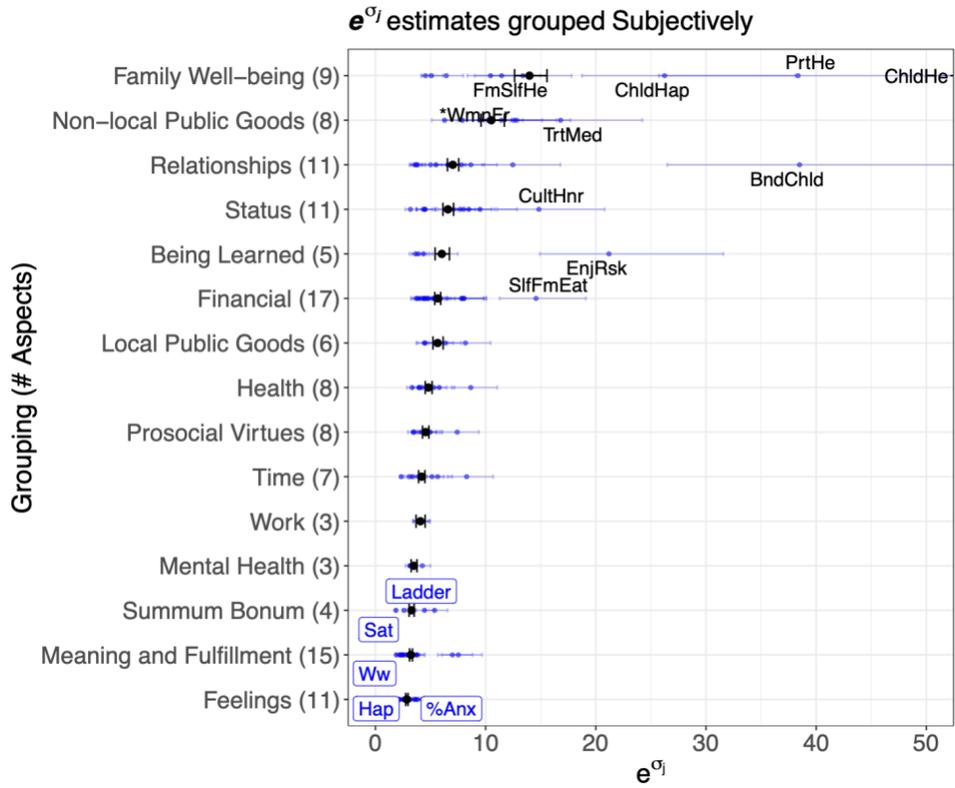e^{σ_j} estimates grouped Subjectively

e^{σ_j} estimates grouped by Embeddings

A3 Persistent vs Transitory variance

One issue in interpreting our estimated preference variation as differences across people is that the it is potentially a combination of true differences across people, and the fact that within-individuals, preferences can change over time.

To account for this, we estimate a modified version of our model. Fix an aspect j, and add a time index t. For now, assume there are only two time periods, $t = 0$ or $t = 1$. We model log marginal utility draws across the two time periods as

$$\begin{bmatrix} m_{ij,1} \\ m_{ij,2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_j \\ \mu_j \end{bmatrix}, \begin{bmatrix} \sigma^2_{m^p} + \sigma^2_{m^\tau} & \cdots \\ \sigma^2_{m^p} & \sigma^2_{m^p} + \sigma^2_{m^\tau} \end{bmatrix} \right)$$

where $\sigma^2{}_{m^p}$ and $\sigma^2{}_{m^\tau}$ are the *persistent* (differences across people) and *transitory* (differences across time) portions of the total variance. To estimate this model, we utilize data coming from two waves of surveys, 3 months apart. In each wave, participants answered tradeoff questions for the same set of 33 aspects. We limit our sample to respondents who answered tradeoffs for all 33 aspects in both waves of the survey. Using this sample, we estimate the results shown in Table A1.

Table A1 – Baseline (33 Aspects): Persistent and Transitory components of variation in preferences
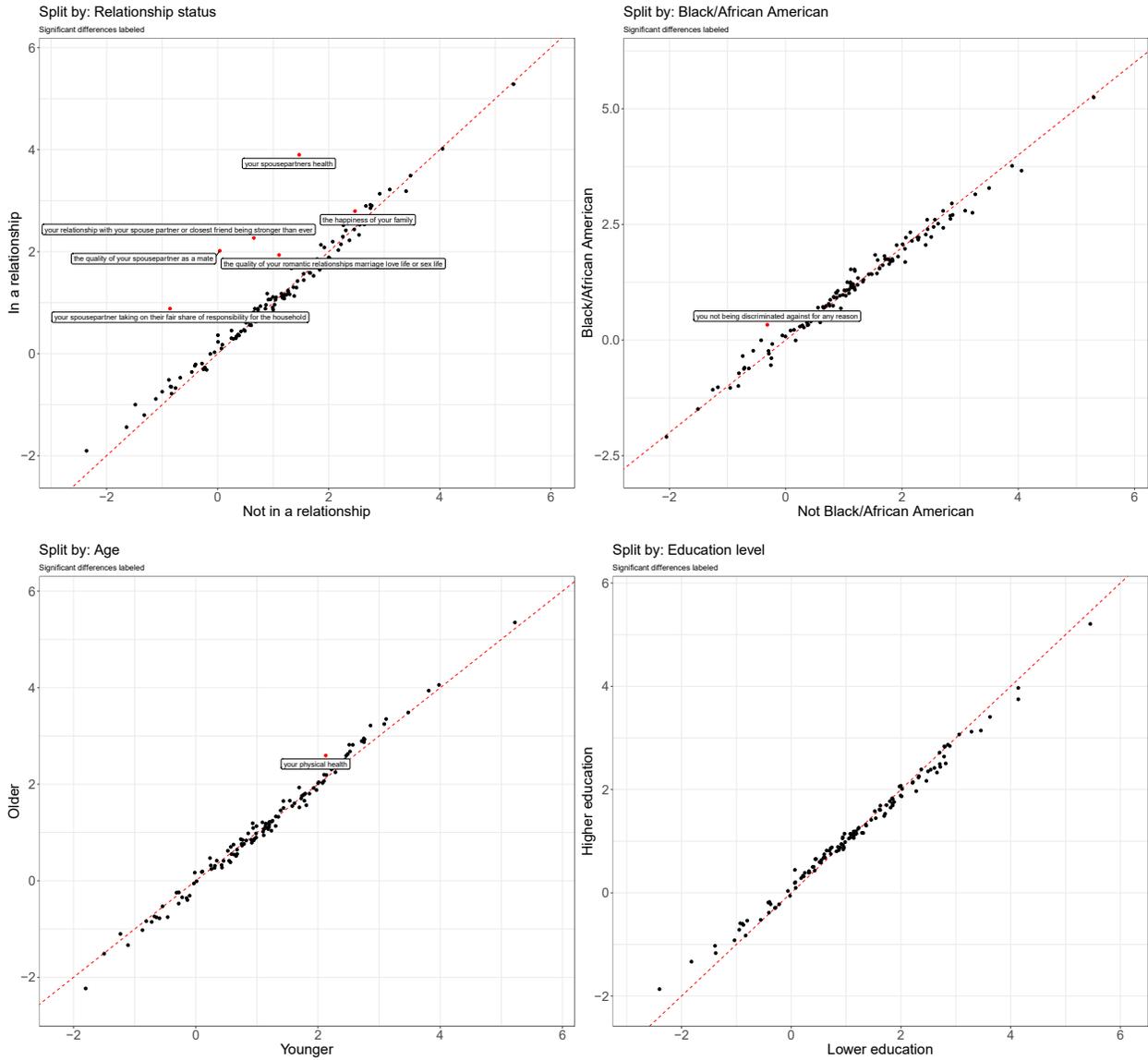
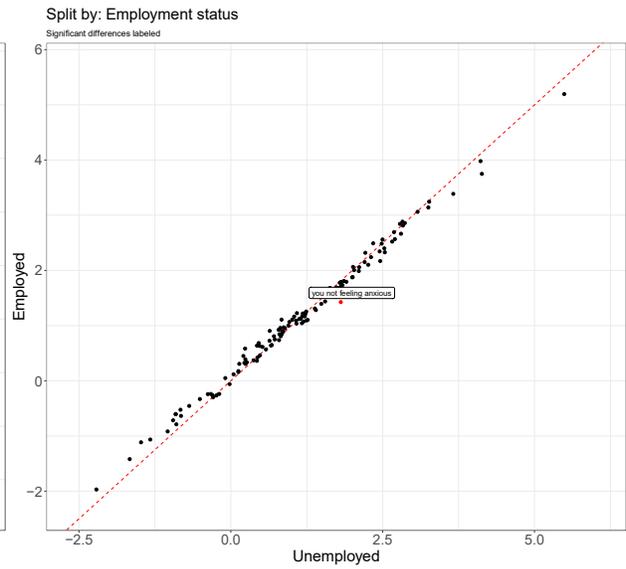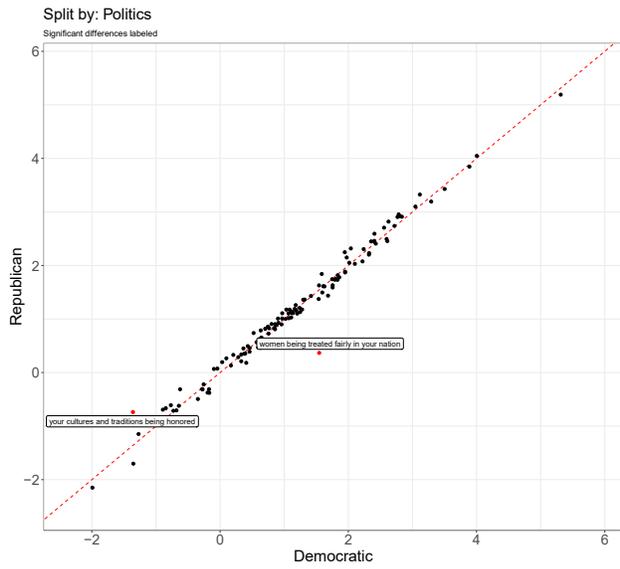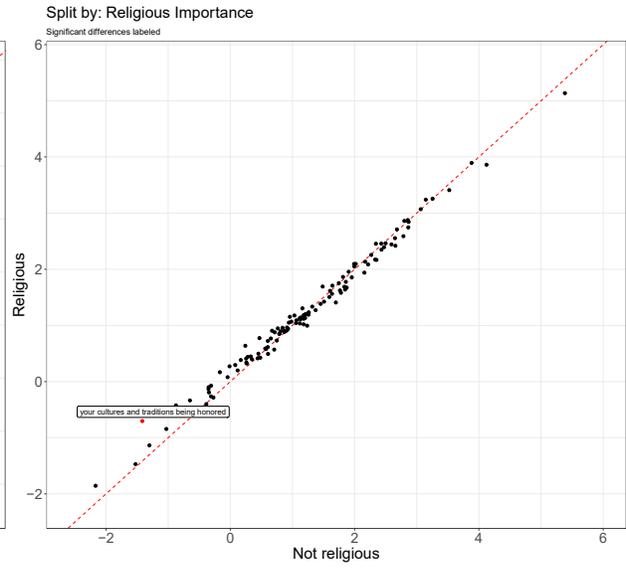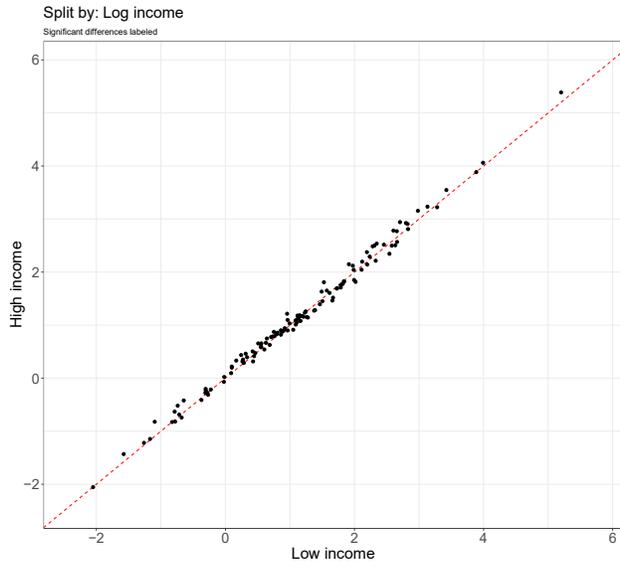|	Comparing pooled and two-wave model	
Parameter	Pooled Estimate (LB - UB)	Two-Wave Estimate (LB - UB)
μ_e	**0.52** (0.49–0.55)	**0.47** (0.44–0.5)
σ_e	**0.77** (0.74–0.8)	**0.7** (0.67–0.73)
σ_m	**1.72** (1.69–1.75)	**1.63** (1.6–1.67)
σ_{mp}	–	**1.07** (1.04–1.09)
σ_{mt}	–	**1.24** (1.21–1.26)

We find that around 60% of the total variance is captured by the transitory component, and 40% is captured by the persistent component, implying that a significant portion of our observed variance is due to differences across time. In column 1, we use the same data, but estimate our

pooled model as described in Section 2.2. Compared to the pooled estimate of total variance, the total implied variance $(\sigma^2{}_{m^p} + \sigma^2{}_{m^\tau})$ is slightly smaller, but of a similar magnitude.

A4 Additional Demographic Splits

Figure A3: Demographic Splits

Split by: Log income
Significant differences labeled

Split by: Religious Importance
Significant differences labeled

your cultures and traditions being honored

Split by: Politics
Significant differences labeled

women being treated fairly in your nation

your cultures and traditions being honored

Split by: Employment status
Significant differences labeled
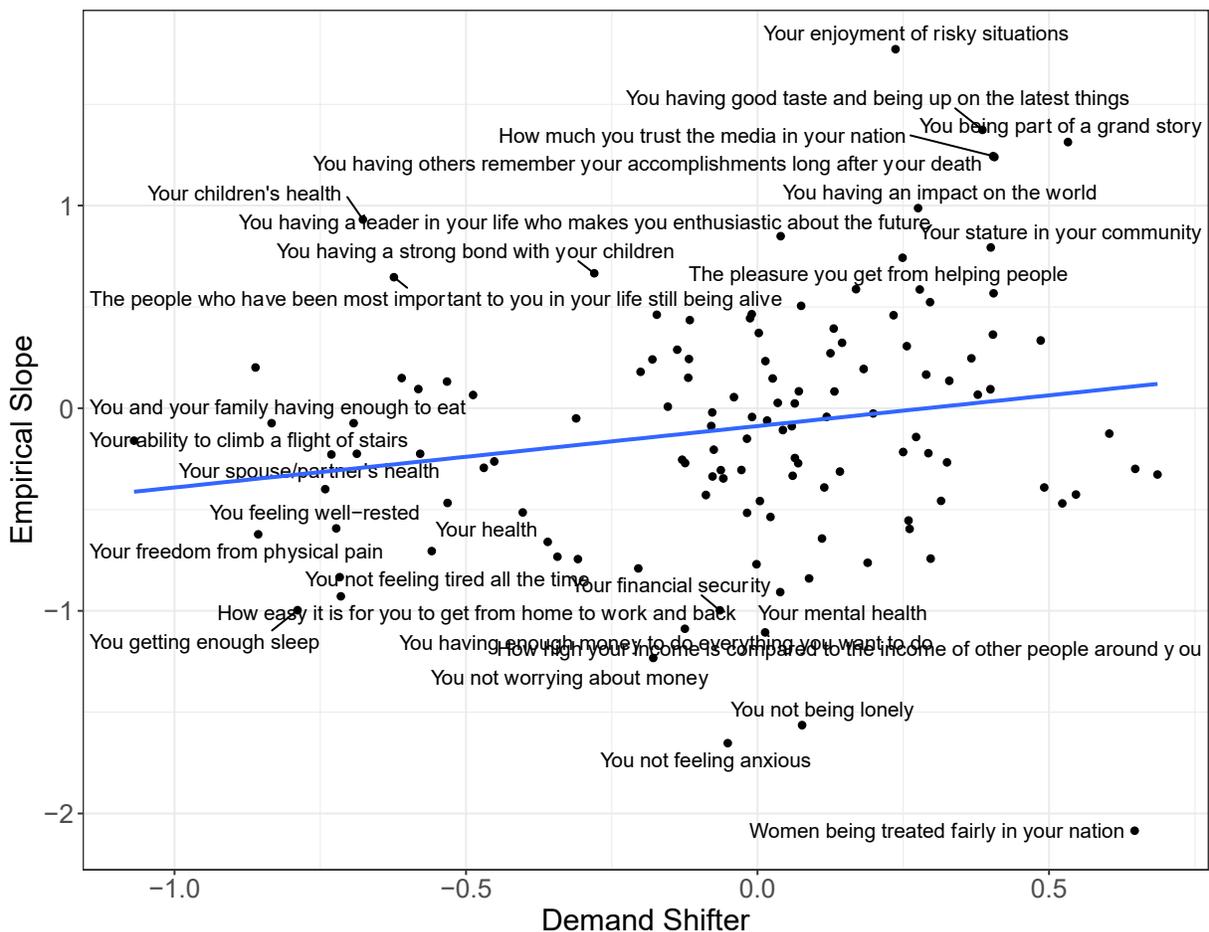
you not feeling anxious

A5 Additional Aspect Flags

Question: People differ in how much they have of each aspect of life. Sometimes it's because of differences in the options available to them, because of differences in resources or abilities. Sometimes it's because of differences in values (how much they prioritize this aspect relative to other aspects of life). Suppose everyone had the same options available to them. How much would people still differ in how much they have of this aspect because of differences in their values?

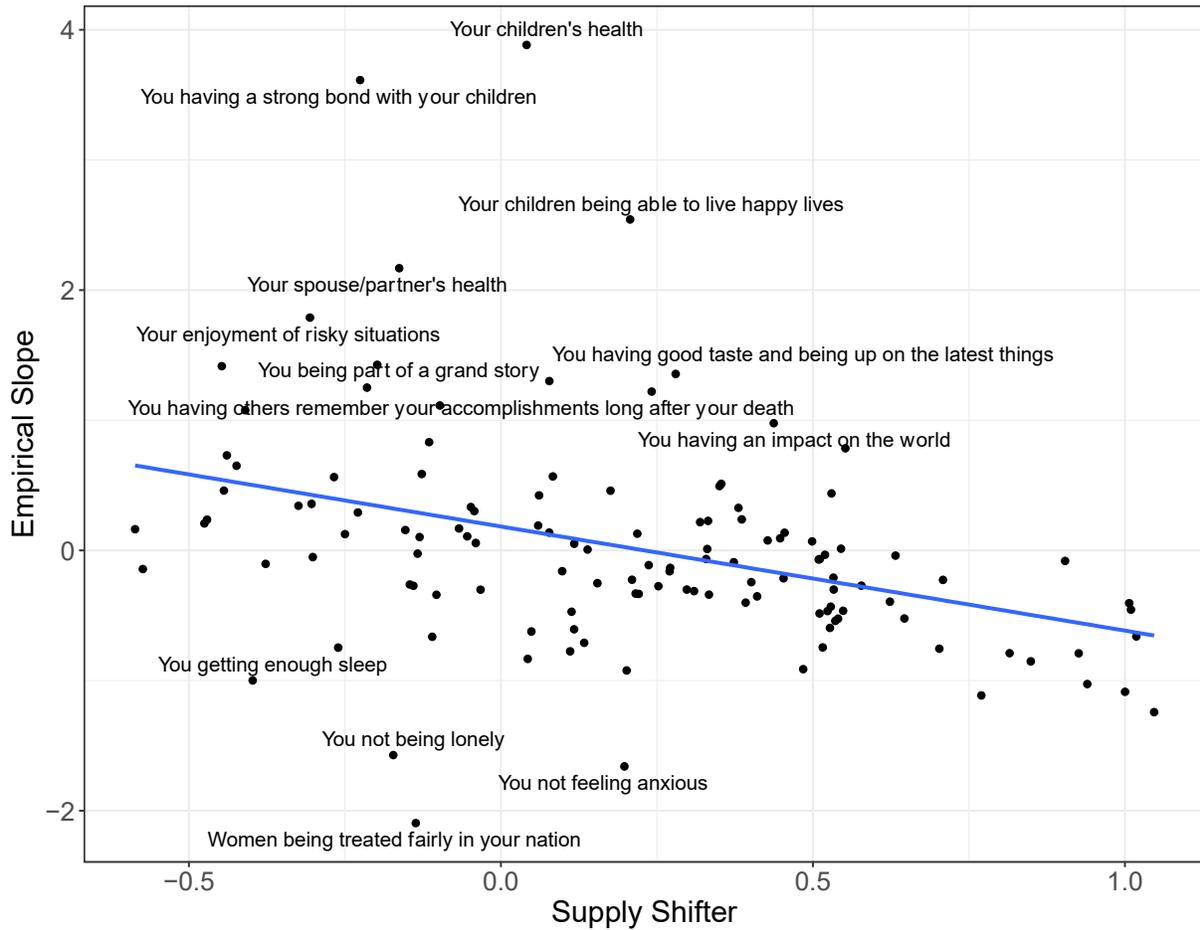Prediction: positive relationship with empirical slope

Figure A4: Empirical Slope vs Demand Shifter Flag

Question: People differ in how much they have of each aspect of life. Sometimes it's because of differences in the options available to them, because of differences in resources or abilities. Sometimes it's because of differences in values (how much they prioritize this aspect relative to other aspects of life). Suppose everyone had the same values. How much would people still differ in how much they have of this aspect because of differences in resources or abilities (and therefore differences in what options are available to them)?

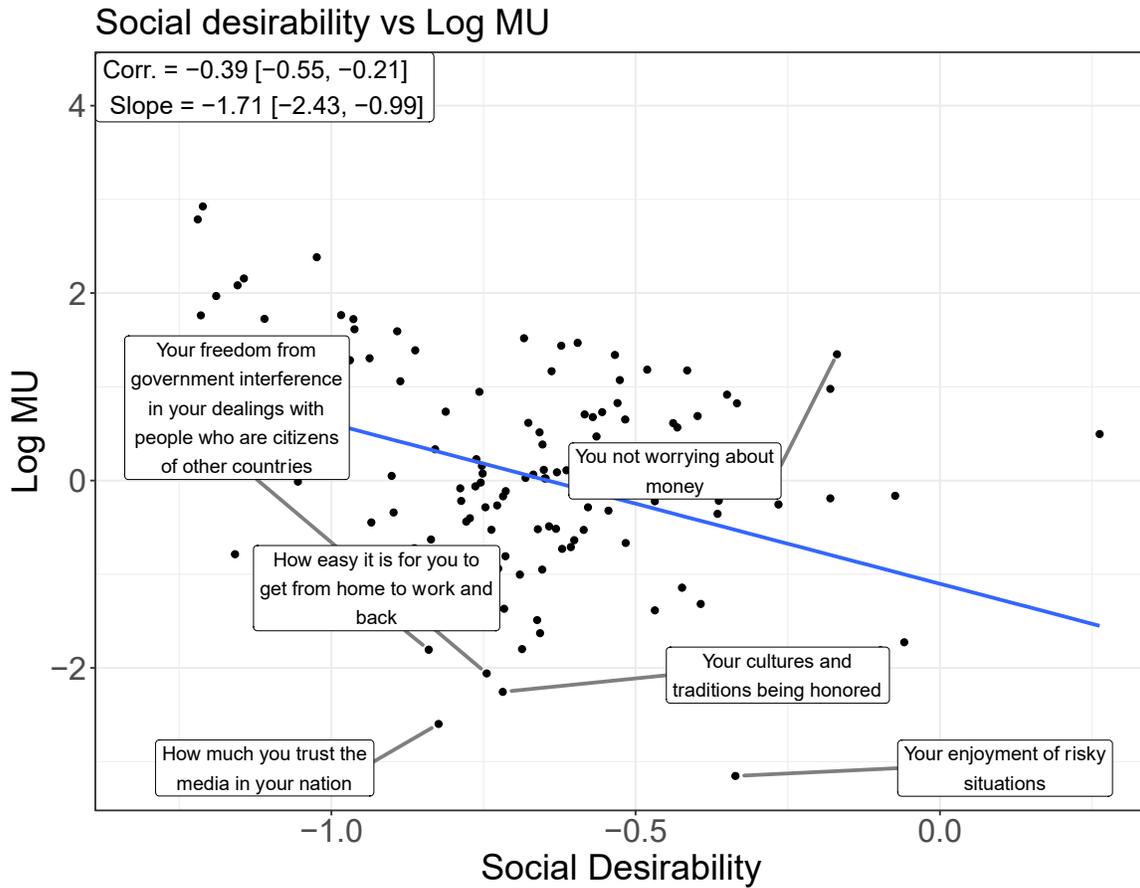Prediction: negative relationship with empirical slope

Figure A5: Empirical Slope vs Supply Shifter Flag

Question: Would someone downplay how much they want this aspect because openly wanting it might make them look bad?

Prediction: Negative relationship with log marginal utility

Figure A6: Log Marginal Utility vs Social Desirability Flag

Social desirability vs Log MU

Corr. = −0.39 [−0.55, −0.21]
Slope = −1.71 [−2.43, −0.99]

Question: How critical is it to you to have at least a little of this aspect in your life rather than none?

Prediction: Much larger MU for lowest rating quintile (or negative relationship with empirical slope)

Figure A7: Empirical Slope vs Inada Condition Flag