

Synthetic personas distort the structure of human belief systems

Christopher Barrie¹, Roberto Cerina²

¹Department of Sociology, New York University

²Institute for Logic, Language, and Computation, University of Amsterdam

February 26, 2026

Abstract

Large language models (LLMs) are increasingly used as synthetic survey respondents, yet it is unclear whether their belief-system structure matches that of real publics. We compare 28 LLMs to the 2024 General Social Survey (GSS) using 52 attitude items and demographic persona traits. We estimate polychoric correlation matrices and propagate uncertainty in the GSS via bootstrap resampling with multiple imputation. Constraint is measured by the variance share explained by the first principal component and by effective dependence, a determinant-based measure of global linear dependence. Across models, LLM personas exhibit substantially higher constraint than humans; conditioning on persona traits reduces constraint far more for LLMs, indicating greater demographic mediation. Projection onto a shared GSS basis further shows overemphasis of the leading dimension and missing secondary structure. These results caution against treating LLM personas as a reliable foundation for synthetic survey data generation.

Significance Statement

Large language models (LLMs) are increasingly used to stand in for survey respondents, often by conditioning on demographic “personas” to simulate diverse publics. That is, practitioners propose prompting an LLM to answer a survey question *as if* it were the description of the person it is given. Comparing responses from a large number of models to a high-quality representative sample of the US population, we show that persona-based prompting produces political belief systems that are unrealistically coherent. Real people commonly hold cross-cutting views that do not fit neatly on a single ideological scale like left or right; LLM personas instead generate answers that are tightly aligned, strongly one-dimensional, and more predictable from demographics. This matters because many applications treat synthetic surveys as substitutes for human samples. Our results suggest that persona-conditioned LLM data can systematically distort the messy character of mass opinion and overstate how consistently demographics map onto ideology.

1 Introduction

A central finding in political behavior, social psychology, and cultural sociology is that mass belief systems are structured but only weakly constrained. Outside of the most engaged subgroups, many citizens hold cross-cutting views; positions on one issue only loosely predict positions on others; and the public is organized along multiple, domain-specific dimensions rather than a single left–right axis [Converse, 2006, Achen, 1975, Zaller, 1992, Boutyline and Vaisey, 2017, Baldassarri and Gelman, 2008]. Demographic attributes predict some attitudes, but they rarely determine whole belief profiles; much of the structure of opinion reflects weak ties among issues, contextual cue-taking, and secondary bundles that are not reducible to demographic sorting [Converse, 2006, Achen, 1975, Zaller, 1992, Boutyline and Vaisey, 2017]. This descriptive fact has a direct measurement implication: representing a public is not only a matter of matching the distribution of answers to individual questions, but also of preserving the relationships among beliefs—which attitudes move together, how many latent dimensions matter, and how strongly demographic attributes organize those di-

mensions. This matters because downstream conclusions in the social sciences hinge on covariance structure rather than on item means alone.

This distinction between marginals and multivariate structure is critical when the “respondents” are generated rather than sampled. Large language models (LLMs) are increasingly used as synthetic survey respondents—“silicon samples”—to approximate public opinion at scale, explore counterfactual publics, and trial experimental interventions [Bisbee et al., 2024, Argyle et al., 2023, Horton, 2023, Li et al., 2025, Hewitt et al., 2024]. The dominant workflow conditions the model on a demographic persona (e.g., age, gender, race, education, income, partisanship) and asks it to answer each item as if it were a respondent with those traits, repeating over many personas to form a synthetic sample [Bisbee et al., 2024, Argyle et al., 2023, Li et al., 2025]. The appeal is obvious—speed, cost, and experimental flexibility—but it raises a basic measurement question: what does it mean for a persona-prompted sample to be “like” a human sample?

Most evaluations target first-order fit: item marginals, subgroup means, or predictive accuracy, with performance that varies by topic and prompting [Bisbee et al., 2024, Santurkar et al., 2023, Wang et al., 2025, Li et al., 2025]. We argue that these tests are incomplete because a model can match marginals yet distort the geometry of belief systems by producing an unrealistically coherent public. This motivates a stricter criterion for synthetic surveys: belief-system structural realism. By structural realism we mean that the relationships among attitudes in synthetic data resemble those in real publics—the pattern and strength of association across items, the effective dimensionality of belief variation, and the degree to which apparent structure is explained by demographic conditioning. This standard mirrors classic construct validity in psychometrics [Cronbach and Meehl, 1955, Campbell and Fiske, 1959] and “utility” in the synthetic microdata literature, where preserving multivariate structure

constitutes the relevant inferential target [Reiter, 2005, Snoke et al., 2018].

We evaluate structural realism by comparing persona-based responses from 28 LLMs to the 2024 General Social Survey (GSS) on a harmonized battery of 52 ordinal attitudes and demographic traits of U.S. residents [NORC at the University of Chicago, 2024]. We estimate latent (polychoric) correlation matrices, propagating uncertainty in the GSS via bootstrap resampling with multiple imputation. We then quantify belief-system constraint (how strongly one attitude predicts others) using two complementary summaries: (i) the share of variance concentrated on the leading principal axis and (ii) a determinant-based index of global linear dependence. Finally, we ask two structural questions that are central to persona prompting: how much of the apparent belief structure is mediated by the persona traits themselves, and which secondary dimensions of human belief structure are missing once we account for demographics.

Across models, persona-conditioned LLM belief systems are systematically over-constrained relative to the human benchmark: attitudes align too tightly, concentrate excessively on a single dominant axis, and become disproportionately predictable from demographic personas. When projected onto a shared human basis, models tend to over-realize the leading human dimension while under-realizing weaker, domain-specific axes that remain present in real publics. The implication is straightforward: even when silicon samples achieve non-trivial first-order agreement, they can still misrepresent the second-order structure of mass opinion in ways that matter for substantive inference.

2 Results

We compare the belief-system structure produced by persona-conditioned large language models (LLMs) to the 2024 General Social Survey (GSS). The GSS is widely regarded as the leading nationally representative social survey in the

United States and serves as a benchmark source for measuring public attitudes [NORC at the University of Chicago, n.d.].

We proceed in three steps. First, we establish that persona-based silicon sampling generates belief systems that are systematically *over-constrained* relative to humans, using both distributional and eigenstructure-based diagnostics. Second, we show that a large share of this apparent constraint is *persona-mediated*: conditioning on demographic persona traits attenuates dominant-axis structure far more in LLMs than in the GSS, consistent with demographic priors carrying too much of the synthetic ideological organization. Third, we identify *which* human dimensions are missing by projecting residual belief structure onto a shared GSS basis and measuring dimension-specific variance deficits. Finally, we relate these structural distortions to conventional accuracy-style benchmarks (squared-weighted Cohen’s κ), which capture agreement in response profiles but do not diagnose whether the underlying correlation geometry is correct (for full details of these analyses, see SI, Figures 12–13).

2.1 Persona-conditioned LLM belief systems exhibit excessive constraint

We begin by assessing whether synthetic personas reproduce the weak-to-moderate ideological constraint that characterizes mass publics. Over-constraint can appear in two complementary ways: (i) as a global inflation of pairwise associations (beliefs becoming too mutually predictable) and (ii) as an over-concentration of variance onto a single latent axis (beliefs collapsing too strongly onto one dominant dimension).

Figure 1 summarizes the distribution of absolute pairwise polychoric correlations using faceted “Saturn” plots, adapted from Tokuda et al. [2025]. For each model and each bootstrap draw, we compute selected quantiles of the absolute

upper-triangular correlations $\{|\rho_{j\ell}|\}$ over all variable pairs, and then take the median of each quantile across bootstrap draws to obtain $\rho_{m,q}$. Each facet plots the 50% probability-mass contour of a bivariate normal distribution with correlation fixed at $\rho_{m,q}$, so tighter ellipses correspond to larger typical dependence at that point in the correlation distribution. Relative to the GSS benchmark (bold black), many LLM belief systems exhibit systematically tighter contours across the mid-quantiles and especially in the upper tail, indicating heavier tails of strong associations and a correlation geometry that is too rigid. To keep the comparison readable with many models, non-highlighted LLMs are drawn as faint grey contour lines, while we color only those model-quantile contours that exceed the GSS by a meaningful margin at that same quantile, $\Delta_{m,q} = \rho_{m,q} - \rho_{\text{GSS},q} \geq 0.10$ (restricted to the largest deviations per quantile).¹

We complement the distributional evidence with eigenvalue-based measures that summarize the severity and shape of global dependence. First, we measure dominant-axis concentration via the share of total system variance explained by the first principal component, $\text{PVE}_1 = \lambda_1/p$. We disaggregate by education level because a core finding in the literature expects that highly educated individuals exhibit more attitudinal constraint [Zaller, 1992]. Figure 2 shows that LLM-generated systems consistently allocate more variance to the leading axis than the GSS. This implies synthetic belief profiles that are excessively one-dimensional: a single latent axis explains too much of what varies across synthetic respondents. The Δ between high- and low-education groups suggests that LLMs do capture this aspect of constraint: higher-educated synthetic individuals exhibit a comparable “constraint premium” to humans, though it is generally attenuated.

Second, we measure global multivariate dependence using effective dependence, $D_e = 1 - |\mathbf{R}|^{1/p}$ [Pena and Rodriguez, 2003]. Unlike PVE_1 , which cap-

¹Complete Saturn plots for all variable pairs are reported in the SI Figure 19.

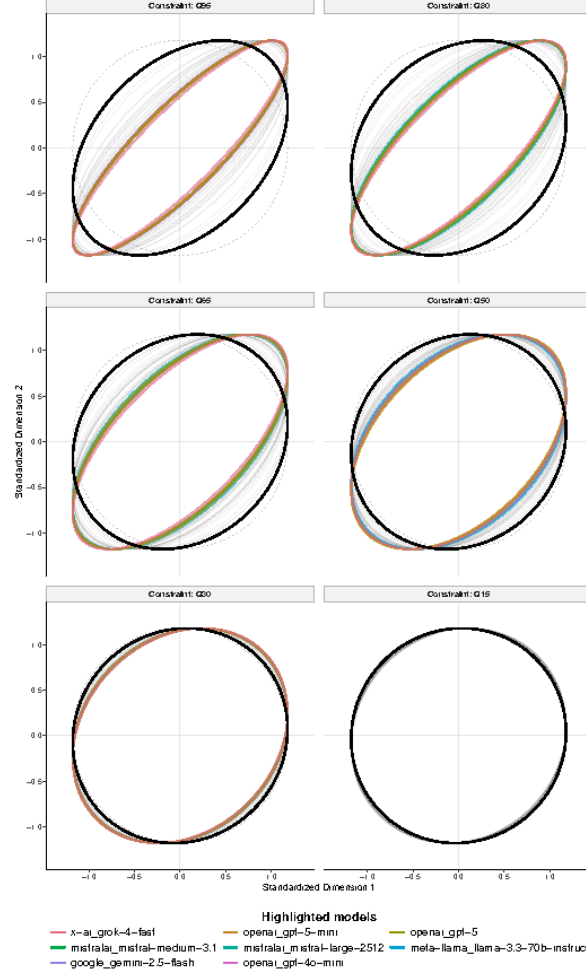


Figure 1: **Faceted Saturn plot summary of system constraint.** Each facet corresponds to a quantile of the absolute correlation distribution $\{|\rho_{j\ell}|\}$ (here $Q_{95}, Q_{80}, Q_{65}, Q_{50}, Q_{30}, Q_{15}$). For each model and quantile q , we compute the q th quantile of $\{|\rho_{j\ell}|\}$ within each bootstrap draw and take the median across bootstraps to obtain $\rho_{m,q}$. We then plot the 50% probability-mass contour of a bivariate normal distribution with correlation set to $\rho_{m,q}$. The dashed circle is the $\rho = 0$ reference contour at the same probability mass. Faint grey contours show all non-highlighted LLMs; the GSS is shown in bold black. Colored contours highlight model-quantile pairs with large positive $\Delta_{m,q} = \rho_{m,q} - \rho_{GSS,q}$ (threshold $\Delta_{m,q} \geq 0.10$, restricted to the largest deviations per quantile), indicating substantially greater constraint than humans at the same point in the correlation distribution.

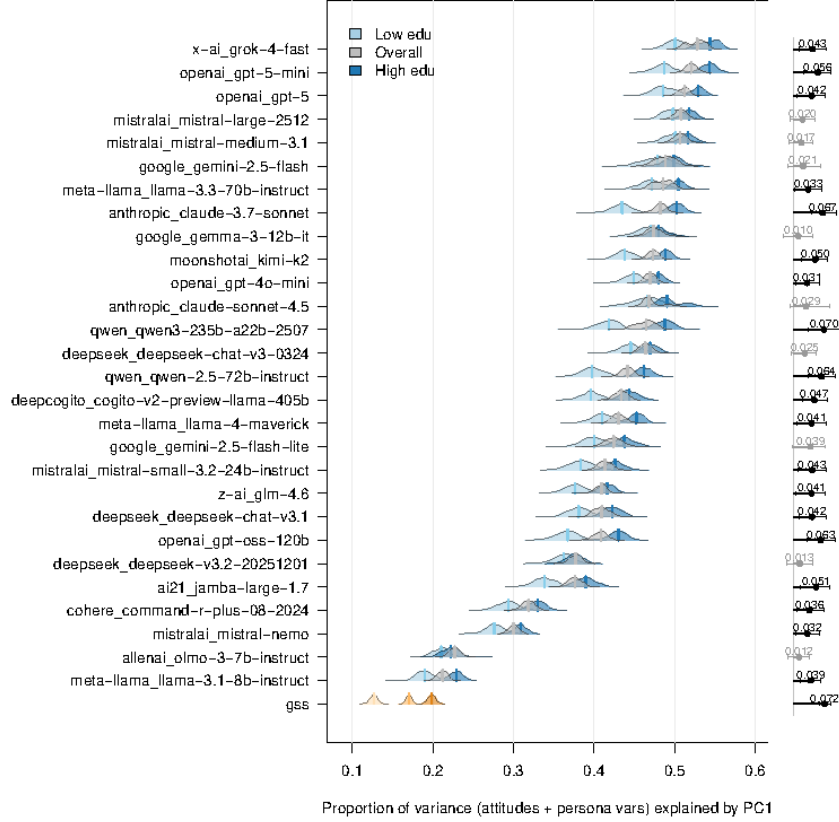


Figure 2: **Dominant-axis constraint by education.** Bootstrap distributions of $PVE_1 = \lambda_1/p$, the proportion of total system variance explained by the first principal component of the full polychoric correlation matrix (attitudes + persona variables), shown for the GSS and each LLM. Distributions are shown overall and stratified by education (low vs. high). Right-side lollipops show the education gradient $\Delta = PVE_1(\text{high edu}) - PVE_1(\text{low edu})$: the point is the median and whiskers denote the 95% bootstrap interval. Lollipops are dark when the 95% interval excludes zero (light otherwise). Larger values indicate more unidimensional belief systems.

tures concentration on the leading axis, D_e is sensitive to the full eigen-spectrum and therefore registers when a system becomes globally near-singular—that is, when variation collapses into a small number of effective dimensions (many eigenvalues shrink toward zero). Figure 3 shows that LLM systems exhibit substantially higher D_e than the human benchmark, indicating implausibly strong global dependence among beliefs and persona traits: responses co-move so tightly that comparatively little independent structure remains. Strikingly, the education gradient in D_e is muted and often statistically indistinguishable from zero for most models, whereas the GSS exhibits a clearer constraint premium among more-educated respondents.

Placing models jointly on the two-dimensional constraint plane makes clear that persona-conditioned silicon samples systematically occupy a region of stronger concentration and stronger global dependence than the GSS benchmark (See SI Figure 8). This gap is large relative to bootstrap uncertainty and persists across education strata, indicating that over-constraint is not confined to a narrow synthetic subgroup.

These results establish that persona-conditioned belief systems are too coherent. We next ask whether this coherence is disproportionately mediated by persona traits—that is, whether demographic conditioning is carrying too much of the synthetic structure.

2.2 Observed constraint is disproportionately mediated by persona traits

Persona-based silicon sampling conditions the model on a demographic profile (e.g., age, gender, race, education, income, party ID), effectively injecting a set of covariates that can act as a strong prior over issue positions. If demographics in synthetic personas are doing too much explanatory work, then conditioning

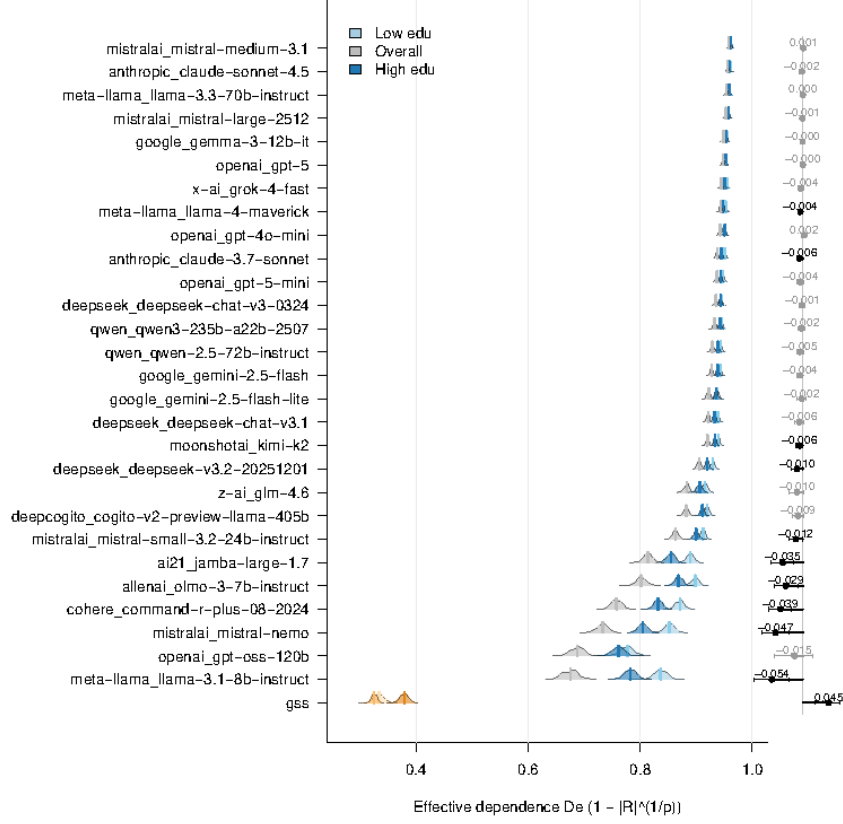


Figure 3: **Global dependence by education.** Bootstrap distributions of effective dependence, $D_e = 1 - |R|^{1/p}$, computed on the full polychoric correlation matrix (attitudes + persona variables) for the GSS and each LLM, shown overall and stratified by education. Right-side lollipops show the education gradient $\Delta = D_e(\text{high edu}) - D_e(\text{low edu})$: the point is the median and whiskers denote the 95% bootstrap interval. Lollipops are dark when the 95% interval excludes zero (light otherwise). Higher D_e indicates stronger global linear dependence in the belief system.

on persona variables should remove a larger fraction of apparent constraint than it does in human survey data.

We test this by computing constraint on the unconditional belief block, $\mathbf{R}_{BB}$, and on the corresponding persona-conditional residual belief structure, $\mathbf{R}_{B|P}$. Figure 4 shows that conditioning yields only modest attenuation of the leading-axis concentration in the GSS, but substantially larger reductions for nearly all LLMs. Substantively, this indicates that a sizable share of the dominant synthetic ideological axis is carried by persona covariation: once demographic traits are partialled out, the residual synthetic belief system becomes much less unidimensional.²

We further decompose cumulative explained variance into persona-mediated and residual components along successive belief principal components. Figure 5 shows that, in the GSS, the cumulative explained variance is dominated by the residual (non-persona) component: most belief-system structure persists after conditioning on demographics. By contrast, in LLM-generated systems the persona-mediated component typically accumulates rapidly and constitutes a large share of the variance captured by early principal components. This pattern is consistent with demographic priors acting as a strong organizing backbone in silicon samples, yielding belief profiles that are more rigidly sorted by persona traits than in human data.³

Persona mediation explains an important part of the over-constraint, but it does not tell us which dimensions of human belief structure are being exaggerated versus erased. We therefore turn to a dimension-specific analysis of “missing structure.”

²For the analogous comparison using D_e rather than PVE_1 , see SI Figure S9.

³A complementary eigen-spectrum diagnostic based on cumulative explained variance (SI Fig. S10) shows the same qualitative pattern: persona-conditioned LLM systems typically reach 90% variance with fewer components than the GSS, consistent with a more low-dimensional belief geometry.

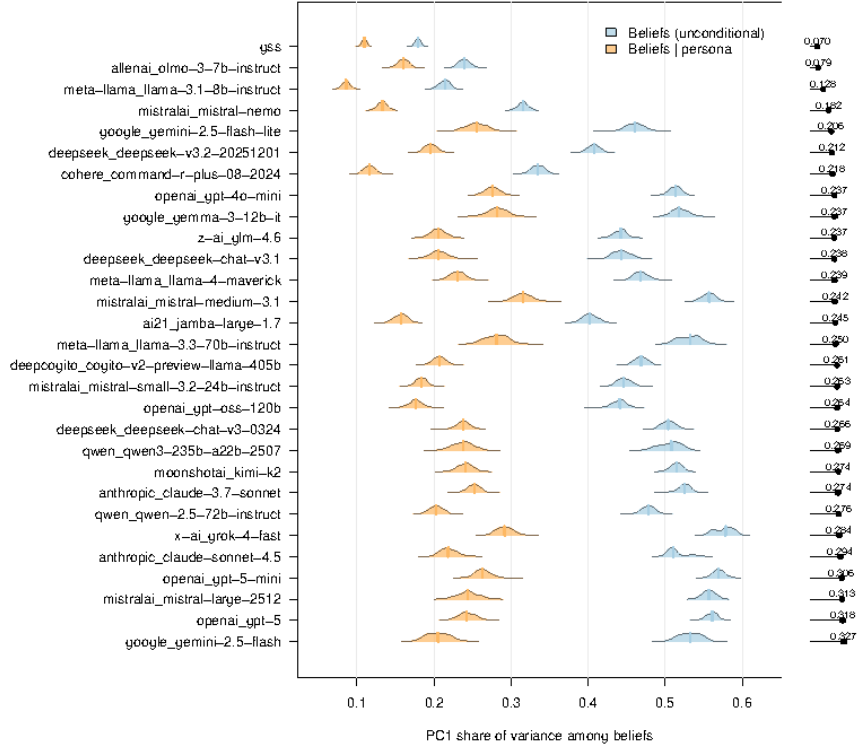


Figure 4: **Persona mediation of dominant-axis belief constraint.** Bootstrap distributions of the proportion of belief variance explained by the first principal component (PVE_1) computed on the unconditional belief block $\mathbf{R}_{BB}$ and on the persona-conditional residual belief matrix $\mathbf{R}_{B|P}$. Right-side lollipops show the median reduction $\Delta PVE_1 = PVE_1(\mathbf{R}_{BB}) - PVE_1(\mathbf{R}_{B|P})$ with 95% bootstrap intervals (whiskers). Lollipops are dark when the 95% interval excludes zero (light otherwise). Larger ΔPVE_1 indicates that the leading ideological axis is more strongly carried by persona-level covariation.

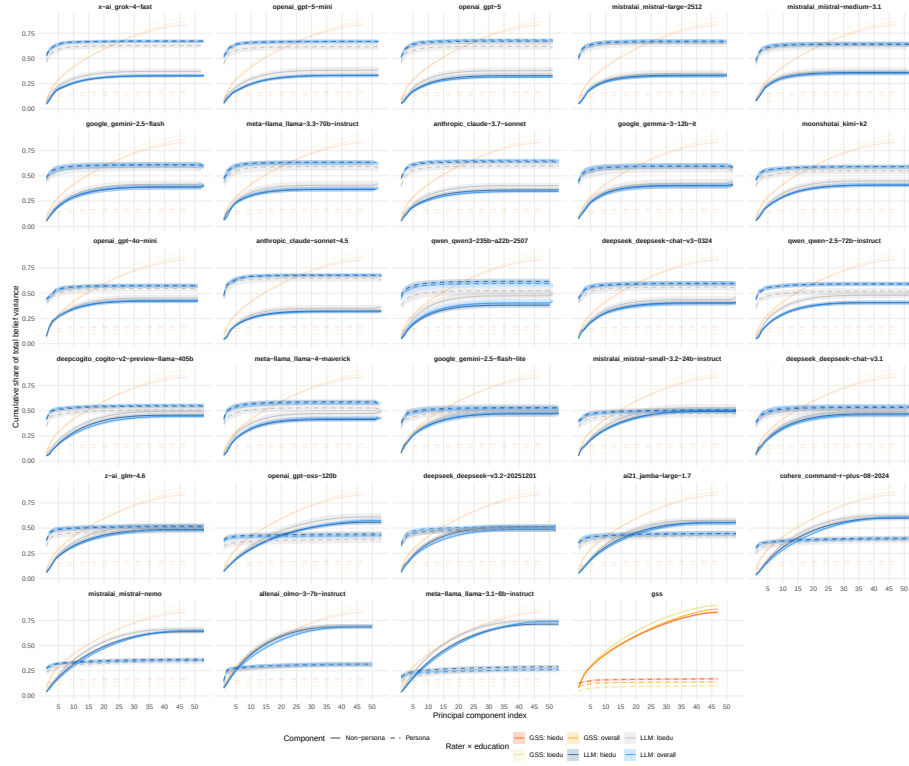


Figure 5: Cumulative persona v. non-persona contributions belief-system variance. For each belief-system model, we decompose the cumulative share of belief variance explained by successive principal components into a persona-mediated component and a residual non-persona component. The x-axis indexes the number of retained belief principal components and the y-axis reports the cumulative share of total belief variance. Colours distinguish overall, low-education, and high-education strata separately for the GSS benchmark and for LLM personas.

2.3 Synthetic personas fail to recreate secondary dimensions of human belief structure

Constraint and persona mediation analyses tell us how much synthetic belief systems differ from humans, but not which parts of human structure are lost. To diagnose missing dimensions, we compare each model to humans on a shared set of axes learned from the GSS residual belief system—that is, the belief correlation matrix after conditioning on persona traits.

Concretely, we first estimate principal components of the GSS residual belief correlation matrix, which provides a set of orthonormal “human” directions in attitude space and the amount of GSS variance associated with each direction. We then ask a simple question for each model: if we project its residual belief system onto the same human directions, how much variance does it exhibit on each one? We compute this projected variance and report it relative to the corresponding GSS benchmark as a Rayleigh ratio.⁴ Ratios near 1 indicate that the model matches the GSS on that dimension; ratios above 1 indicate that the model amplifies that human dimension; and ratios below 1 indicate that the model attenuates it.

Figure 6 shows a consistent pattern across models: many LLM personas over-realize the leading GSS dimension, while under-realizing multiple higher-order components that remain present in the human benchmark even after controlling for persona traits. This implies that silicon samples are not merely “too ideological” in the sense of stronger overall alignment; rather, they disproportionately concentrate variance on the dominant axis while compressing or erasing weaker, domain-specific and cross-cutting dimensions that contribute to the multidimensional geometry of real belief systems. The GSS loading patterns used to interpret these axes are reported in the SI (Figure S11).

⁴See Data and Method section and SI for formal notation and description of these calculations.

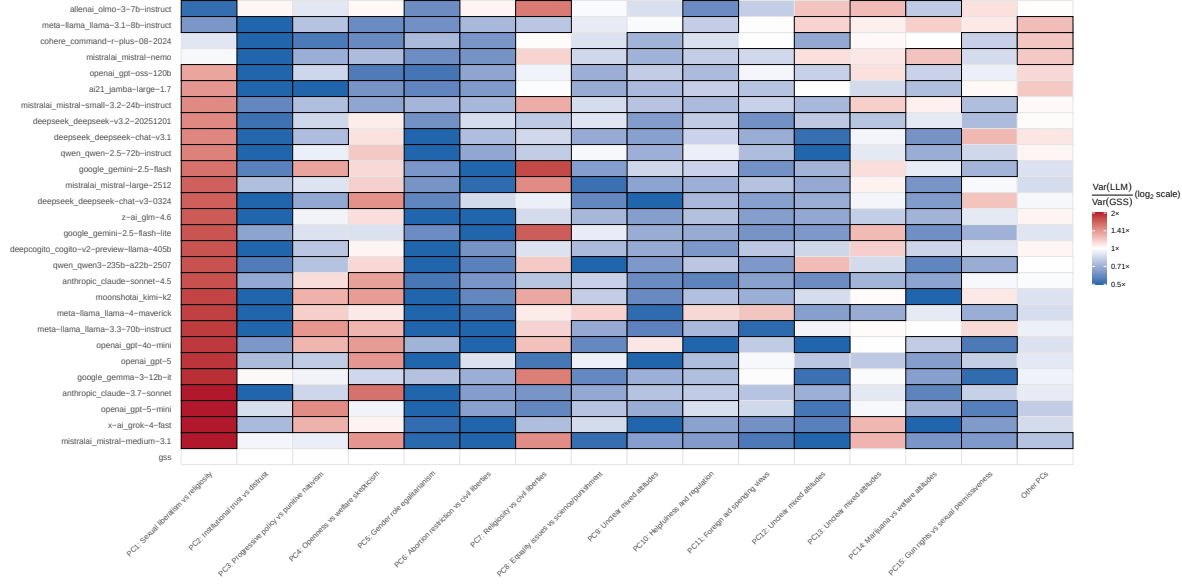


Figure 6: **Missing structure via Rayleigh ratios on a shared GSS basis.** Each cell reports the Rayleigh ratio $RR_{LLM,j} = v_{LLM,j}/\lambda_j$, where $v_{LLM,j} = \mathbf{u}_j^\top \mathbf{R}_{LLM} \mathbf{u}_j$ is the variance of a model’s persona-conditional belief system projected onto GSS principal-component direction $\mathbf{u}_j$, and λ_j is the corresponding GSS eigenvalue. Fill encodes $\log_2(RR_{LLM,j})$ (white = $1\times$; blue $< 1\times$; red $> 1\times$), truncated to the display range. We display the leading GSS components and pool the remaining tail of higher-order PCs into an “Other PCs” column using variance-share weighting. Cell opacity increases with the number of usable bootstrap draws (n_{boot}). Black outlines indicate components whose 95% interval for $RR_{LLM,j}$ excludes 1. Models are ordered by over-realization of the leading GSS component (PC1). Other PCs aggregates remaining components and should be interpreted as a summary of tail variance, not a coherent ideological axis.

2.4 Structural realism and conventional agreement benchmarks

A natural question is whether these structural distortions are visible to conventional evaluation metrics that focus on response-level agreement. Using squared-weighted Cohen’s κ to quantify net-of-chance agreement between each model’s response matrix and the GSS, we find that LLM-GSS agreement is typically positive but systematically weaker than LLM-LLM agreement, and it varies meaningfully across models (SI Figure 12). Moreover, when we relate $\kappa_{\text{LLM,GSS}}$ to the constraint diagnostics, more constrained models often show higher agreement on the dominant human dimensions (SI Figure 13). This pattern underscores why first-order “accuracy” checks are incomplete: a model can improve agreement with the GSS in aggregate while still collapsing belief variance onto a dominant axis, inflating global dependence, and under-realizing the weaker secondary dimensions that govern important relational inferences about belief-system structure.

3 Conclusion

In this study, we find that, across 28 models, persona-conditioned silicon samples are markedly over-constrained relative to relevant human benchmarks: correlations among items are inflated, variance collapses too strongly onto a single dominant axis, and much of the apparent ideological structure is disproportionately mediated by persona traits. When we compare residual belief structure to a shared human basis, models tend to amplify the leading human dimension while attenuating the weaker, domain-specific axes that give real publics their characteristically multidimensional, cross-cutting structure. These distortions are not reliably detected by conventional agreement-style benchmarks alone,

underscoring that “accuracy” on marginals or response profiles is not equivalent to structural realism.

The implication is not that LLMs are useless as survey substitutes, but that persona-based silicon sampling exhibits identifiable structural markers that matter for downstream inference. For applications where substantive claims depend on covariance geometry—polarization, ideological dimensionality, belief-network modularity, or the degree to which demographics explain attitudes—over-constrained synthetic data can yield confident but misleading inferences, including exaggerated ideological coherence and overstated demographic determinism. At minimum, evaluations of synthetic respondents should therefore include second-order diagnostics alongside first-order fit, and should explicitly test whether the role of demographic covariates in organizing beliefs matches human benchmarks.

We hope that our analyses provide a template for these diagnostic tests. We believe that an emerging agenda should focus on prompting, calibration, and post-processing methods that target structural quantities as estimands, not merely marginal distributions. In the near term, structural checks like the Saturn-plot quantiles, eigen-spectrum diagnostics, and persona-conditional residualization provide practical tools for auditing silicon samples before they are used for substantive inference. Until synthetic respondents can reliably preserve the weak, heterogeneous, and partially cross-cutting structure of real belief systems, they should complement—not replace—human survey data in studies of mass opinion.

4 Materials and Methods

Data and synthetic personas. We benchmarked synthetic respondents against the 2024 General Social Survey (GSS), a flagship repeated cross-sectional survey

of U.S. adults [Davern et al., 2024]. We sampled 1,000 GSS respondents to seed synthetic personas: each persona was a natural-language biography generated from the respondent’s coded attributes (age, sex, race, marital status, education, family income, party identification, political ideology, religious affiliation, and religious service attendance). Missing attributes were omitted. This procedure follows the “silicon sampling” paradigm in which LLMs are prompted to role-play demographically grounded profiles [Argyle et al., 2023, Park et al., 2024, Bisbee et al., 2024]. Details of the full prompt design are in SI Figures 30 and 31.

Survey instrument. Each persona was administered 52 ordinal attitudinal items drawn from the GSS. Item selection followed prior work operationalizing belief-system constraint using GSS batteries [DellaPosta, 2020]. Topics thus spanned culture-war topics and non-culture-war domains (public spending priorities, institutional confidence, social trust, and economic outlook). Response scales ranged from binary to 7-point ordered categories. Full question wordings and response options are provided in the SI.

LLM querying. We queried 28 large language models (SI Table 1) via the OpenRouter API. For each model-persona-item, the prompt specified (i) temporal context (“It is now 2024”), (ii) the persona biography (“You are answering as the following person living in the United States:”), (iii) the item text and numbered response options, and (iv) an instruction to return only the number of the chosen option. Temperature was set to 1.0 and each persona-item pair was queried once per model. Non-parseable outputs (including non-numeric responses or out-of-range option numbers) were coded missing.

Harmonization and recoding. Human GSS responses and LLM outputs were harmonized to a common long format (persona $\times$ item $\times$ response). “Don’t

know,” “refused,” and analogous non-substantive codes in the GSS were recoded as missing. Persona attributes were recoded as ordinal variables for joint correlation analysis: age was binned into six categories; sex, race, and marital status were dichotomized (female, White, married); income was collapsed to eight ordered categories; party identification retained a 7-point scale (excluding “other party”). Education strata were defined as low (no college; ≤ 13 years) versus high (some college or more; ≥ 14 years).

Polychoric correlations, imputation, and bootstrap uncertainty. All variables were treated as ordinal and association structure was summarized using polychoric correlation matrices, estimated via the two-step procedure [Olsson, 1979] as implemented in `lavaan` [Rosseel, 2012]. To propagate uncertainty from sampling variation and missing data in the human benchmark, we combined bootstrap resampling with multiple imputation by chained equations (MICE) using random forests (`miceRanger`) [van Buuren and Groothuis-Oudshoorn, 2011, Wilson, 2020]. Specifically, we generated $B = 500$ bootstrap resamples at the persona (respondent) level, and for each resample completed missing values by drawing from one of $m = 30$ imputed datasets. Correlation matrices were eigenvalue-smoothed to ensure positive semidefiniteness. The same pipeline was applied overall and within education strata. Additional diagnostics on imputation stability are reported in the SI.

Constraint and persona mediation. From each bootstrap correlation matrix $\mathbf{R} \in \mathbb{R}^{p \times p}$ we computed two constraint diagnostics: (i) dominant-axis concentration $\text{PVE}_1 = \lambda_1/p$, where λ_1 is the largest eigenvalue [Jolliffe and Cadima, 2016]; and (ii) effective dependence $D_e(\mathbf{R}) = 1 - |\mathbf{R}|^{1/p}$, a determinant-based measure of global linear dependence [Pena and Rodriguez, 2003]. To quantify the extent to which apparent belief constraint is mediated by persona at-

tributes, we partitioned $\mathbf{R}$ into belief and persona blocks and computed the persona-conditional residual belief correlation matrix via a Schur complement, then recomputed PVE_1 and D_e on the residual belief structure.

Missing-structure projections. To identify which human belief dimensions are under- or over-realized by synthetic personas, we constructed a GSS-derived orthonormal basis from the persona-conditional belief correlation matrix and projected each model’s persona-conditional belief system onto this basis using Rayleigh quotients. We report Rayleigh ratios comparing model variance to GSS variance along each GSS principal direction.

Agreement with the GSS benchmark. To relate structural realism to conventional accuracy metrics, we computed quadratically weighted Cohen’s κ for ordered categories between each model’s persona–item responses and the GSS benchmark over the common set of non-missing cases [McHugh, 2012]. Pairwise agreement was used descriptively to summarize first-order similarity, and was not used to define constraint.

Software and Reproducibility

All materials to reproduce analyses will be made available upon publication.

Acknowledgments

We thank audiences at the Meta Computational Social Science Seminar, the NYU Data Science Frontiers: Society and Politics Conference, the NYU Center for Social Media, AI and Politics Research Seminar, the Columbia Political Methodology Colloquium, and the Toronto Data Workshop for useful feedback.

References

- Philip E Converse. The nature of belief systems in mass publics (1964). *Critical review*, 18(1-3):1–74, 2006.
- Christopher H. Achen. Mass political attitudes and the survey response. *American Political Science Review*, 69(4):1218–1231, 1975.
- John Zaller. *The Nature and Origins of Mass Opinion*. Cambridge University Press, Cambridge, 1992.
- Andrei Boutyline and Stephen Vaisey. Belief network analysis: A relational approach to understanding the structure of attitudes. *American journal of sociology*, 122(5):1371–1447, 2017.
- Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446, 2008.
- James Bisbee, Joshua D Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M Larson. Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, 32(4):401–416, 2024.
- Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.
- John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus?, April 2023. URL <http://www.nber.org/papers/w31122>.
- Ang Li, Haozhe Chen, Hongseok Namkoong, and Tianyi Peng. Llm generated persona is a promise with a catch. *arXiv preprint arXiv:2503.16527*, 2025.

- Luke Hewitt, Ashwini Ashokkumar, Isaias Ghezze, and Robb Willer. Predicting results of social science experiments using large language models. *Preprint*, 2024.
- Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. Whose opinions do language models reflect? In *International Conference on Machine Learning*, pages 29971–30004. PMLR, 2023.
- Angelina Wang, Jamie Morgenstern, and John P Dickerson. Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, 7(3):400–411, 2025.
- Lee J. Cronbach and Paul E. Meehl. Construct validity in psychological tests. *Psychological Bulletin*, 52(4):281–302, 1955.
- Donald T. Campbell and Donald W. Fiske. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 56(2): 81–105, 1959.
- Jerome P Reiter. Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 168(1):185–205, 2005.
- Joshua Snoke, Gillian M Raab, Beata Nowok, Chris Dibben, and Aleksandra Slavkovic. General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(3):663–688, 2018.
- NORC at the University of Chicago. General social survey 2024 cross-section, 2024.
- NORC at the University of Chicago. About the GSS. <https://gss.norc.org/about-the-gss.html>, n.d. Accessed 2026-01-29.

- Tomoki Tokuda, Ben Goodrich, Iven Van Mechelen, Andrew Gelman, and Francis Tuerlinckx. Visualizing distributions of covariance matrices. *Journal of Data Science, Statistics, and Visualisation*, 5(7), 2025.
- Daniel Pena and Julio Rodriguez. Descriptive measures of multivariate scatter and linear dependence. *Journal of Multivariate Analysis*, 85(2):361–374, 2003.
- Michael Davern, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. General social survey, 2024, 2024. URL <https://doi.org/10.17605/OSF.IO/M6K2J>. Data set.
- Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*, 2024.
- Daniel DellaPosta. Pluralistic collapse: The “oil spill” model of mass opinion polarization. *American Sociological Review*, 85(3):507–536, 2020. doi: 10.1177/0003122420922989. URL <https://doi.org/10.1177/0003122420922989>.
- Ulf Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460, 1979.
- Yves Rosseel. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36, 2012. doi: 10.18637/jss.v048.i02.
- Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67, 2011. doi: 10.18637/jss.v045.i03.
- Sam Wilson. miceranger: Multiple imputation by chained equations with random forests. *R package version*, 1(5), 2020.

- Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.
- David B Flora and Patrick J Curran. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9(4):466–491, 2004.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer series in statistics New-York, 2009.
- Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.
- Jonathan Goodman. Class 6, linear problems, more. Lecture notes for *Stochastic Calculus* (MATH-GA 2902.001), Courant Institute of Mathematical Sciences, New York University, September 2019. URL <https://math.nyu.edu/~goodman/teaching/StochCalc2019/notes/Class6.pdf>. Fall 2019 course notes.
- John William Strutt Baron Rayleigh. *The theory of sound*, volume 2. Macmillan, 1896.
- Matthias Gamer, Jim Lemon, Ian Fellows, and Puspendra Singh. *irr: Various Coefficients of Interrater Reliability and Agreement*, 2019. URL <https://CRAN.R-project.org/package=irr>. R package version 0.84.1.
- Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

Synthetic personas distort the structure of human belief systems

Supporting Materials (SM)

A Estimating Polychoric Correlations

Beliefs and persona-variables are often assessed using Likert-type response scales, which are ordinal and only imperfectly reflect an underlying continuum. Computing Pearson correlations between item scores assumes that equal steps on the response scale correspond to equal steps on the latent belief-persona continuum and that the marginal distributions of the individual Likert items do not distort the association. But when the scales are low-dimensional or the response distributions are highly unbalanced, these assumptions break down: chance agreement and skewed marginals can inflate or deflate the observed correlation. A more appropriate approach is therefore to model the association at the latent level net of the marginals using polychoric correlations [Olsson, 1979].

Formally, let $\mathcal{B}$ index the o beliefs and $\mathcal{P}$ index the q persona variables, and define $p := o + q$. Each of the p variables is represented by a K_j -point Likert scales, or another equivalent ordinal representation; let $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})^\top$, where $y_{ij} \in \{1, \dots, K_j\}$ for $j = 1, \dots, p$, denote the observed ordinal responses for individual i ; let $y_{ij} = k \iff \tau_{j,k-1} < y_{ij}^* \leq \tau_{j,k}$, where $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{ip}^*)^\top$ are the set of latent propensities, and $\boldsymbol{\tau}_j$ contains thresholds indicating the level of latent propensity at which individual i is compelled to select the next point on the ordinal scale, such that $-\infty = \tau_{j,0} < \dots < \tau_{j,K_j} = +\infty$. The underlying latent propensities are assumed to follow a multivariate normal model:

$$\mathbf{y}_i^* \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad R_{jj} = 1, \quad R_{j\ell} = \rho_{j\ell} \quad (j \neq \ell). \quad (1)$$

A drawback of polychoric correlations is that they can be difficult to estimate reliably: because they are obtained by maximizing a likelihood over threshold and correlation parameters under a latent bivariate normal model, the numerical optimization can become ill-conditioned (e.g., with sparse contingency tables,

extreme or highly unbalanced categories, or near-perfect associations), leading to slow convergence, nonconvergence, or estimates that are sensitive to small perturbations in the data.⁵ We are particularly sensitive to issues around near-perfect associations, which are a feature in LLMs, as we discuss below.

To address issues around optimisation we opt for a *two-stage* estimator [Olson, 1979]: thresholds are first estimated from the univariate marginal distributions of each ordinal item, and then (holding those thresholds fixed) the polychoric correlations are estimated from the bivariate contingency tables. The trade-off is that two-step estimation is generally significantly faster and more stable, though less statistically efficient than full-information maximum likelihood, and it can inherit bias or instability when the univariate threshold estimates are poor (again most likely with very sparse or extreme marginals). We expect two-stage estimates to be relatively robust for our purposes [Flora and Curran, 2004], though we acknowledge our LLM estimates will typically be biased due to high multi-collinearity between items. Generally, the resulting bias is conservative with respect to the system-constraint derived from the polychoric correlation matrices of LLMs.

We implement this via the `lavaan` [Rosseel, 2012] package in R. The package implements the two-stage estimator in pairwise fashion, populating a $p \times p$ correlation matrix with bivariate correlations, and enforcing positive-semidefiniteness via thresholding eigenvalues.

A.1 Propagating Uncertainty via Bootstrap & MICE

To account for uncertainty we generate 500 bootstrap samples [Hastie et al., 2009] of the GSS, and estimate a polychoric correlation matrix for each sample, obtaining what are in effect an approximate set of Bayesian posterior samples.

⁵We write $\mathbf{R}$ for a generic latent correlation matrix. In theory, $\mathbf{R}$ denotes the population latent correlation matrix; in practice, we evaluate all downstream quantities by plug-in using an estimated polychoric matrix.

A challenge with GSS data is represented by missing values. After recoding ‘*don’t know*’s and ‘*prefer not to say*’ as NAs, and some additional data wrangling, missing values remain widespread. This is simply because not all items are asked of all respondents. To address this we rely on Multiple Imputation via Chained Equations (MICE) [van Buuren and Groothuis-Oudshoorn, 2011]. We leverage a set of chained random forest models via `miceRanger` [Wilson, 2020] to integrate the imputation procedure with the bootstrapping. First, we learn a set of 30 (or any ‘large-enough’ number) plausible imputation models via random initialisation of the chained equations, representing various data generating processes (DGPs) for the missing values which are compatible with the data at hand. We follow by sampling candidate values to fill the NAs via predictive mean matching, selecting a DGP at random amongst those we learned above, for each bootstrap sample. This results in complete bootstrap samples with realistic levels of variability, which further preserve the original association between variables on average (see Figure 14). Figure 7 presents the resulting polychoric correlation matrix from the GSS. Let $\tilde{\rho}_{j\ell} := \text{med}_{b=1,\dots,B} \{\rho_{j\ell}^{(b)}\}$, then the following is a measure of statistical significance of the direction of the correlation over the bootstrap simulations, representing the proportion of instances in which a given pairwise correlation for bootstrap sample b , $\rho_{j\ell}^{(b)}$, has the same sign as the median correlation

$$\Pr(\text{sign})_{j\ell} = \frac{1}{B} \sum_{b=1}^B \left[\text{sign}(\rho_{j\ell}^{(b)}) = \text{sign}(\tilde{\rho}_{j\ell}) \right].$$

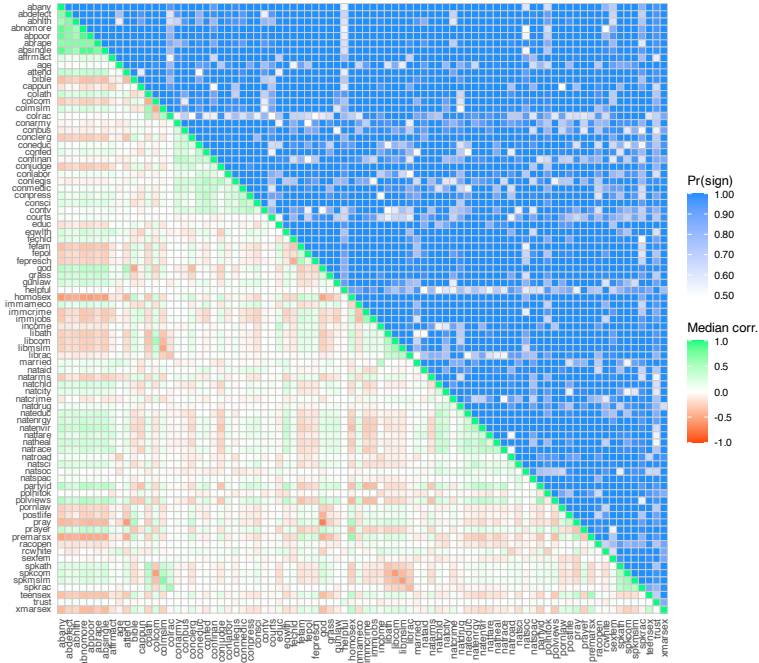


Figure 7: **Polychoric correlation structure in the 2024 GSS.** The lower triangle reports the median polychoric correlation across bootstrap–imputation replications for all attitude items and persona variables. The upper triangle reports sign stability, $\text{Pr}(\text{sign})_{j\ell}$, the proportion of bootstrap replications in which the estimated correlation has the same sign as the median correlation (color scale: 0.5 to 1). Together, the two triangles summarise both the typical association strength and the robustness of its direction under resampling and imputation.

B Visualising System Constraints: *Saturn* Plots

Following Tokuda et al. [2025], we visualise polychoric correlations by mapping a target correlation ρ to the 50% probability-mass contour of a standardized bivariate normal distribution with correlation ρ . This yields an intuitive geometric depiction of pairwise dependence: as $|\rho|$ increases, the contour becomes a narrower ellipse (degenerating to a line as $\rho \rightarrow \pm 1$), while as $\rho \rightarrow 0$ it approaches the $\rho = 0$ reference circle.

Figure 19 reports Saturn contours for the full set of variable pairs.⁶ A common descriptive summary of a $p \times p$ correlation matrix is the mean absolute correlation (MAC),

$$\text{MAC}(\mathbf{R}) = \frac{2}{p(p-1)} \sum_{1 \leq j < \ell \leq p} |\rho_{j\ell}|,$$

which accords with the intuition that more constrained systems tend to exhibit larger typical $|\rho|$. However, MAC is a poor *quantitative* measure of system constraint for at least four reasons: (i) it inherits the well-known positive bias of $|\hat{\rho}|$ under sampling variation, an effect that compounds as the number of pairs grows; (ii) it is tail-sensitive, so a small number of large correlations can dominate the mean; (iii) it discards sign structure and dimensional organization by collapsing $\rho_{j\ell}$ to $|\rho_{j\ell}|$ and then to a single average; and (iv) it is inflated by item redundancy/local dependence when multiple variables tap the same latent construct. Thus, while constrained systems often have high MAC, high MAC is neither necessary nor sufficient for strong constraint.

For these reasons, we use Saturn plots as *distributional* diagnostics rather than as visualizations of MAC alone: for each rater we summarize the distribution of $\{|\rho_{j\ell}|\}$ by selected quantiles and compare these quantiles to the GSS

⁶If desired, we can also provide a reduced version that omits pairs involving persona variables.

benchmark via $\Delta_{m,q}$ (displayed in separate facets).

C Quantifying Severity and Structure of Constraints

Eigenvalue Decomposition (EVD) [Hastie et al., 2009, Jolliffe, 2011] provides a unified framework to analyse the degree of underlying dependence in the system of beliefs. Recall that our belief-system object is a symmetric correlation matrix $\mathbf{R} \in \mathbb{R}^{p \times p}$, from which it follows that:

$$\mathbf{R} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top, \quad \mathbf{U}^\top \mathbf{U} = \mathbf{I}_p, \quad \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p), \quad \lambda_1 \geq \dots \geq \lambda_p, \quad (2)$$

where columns $\mathbf{u}_j$ of $\mathbf{U}$ define the j th principal-component *direction* (also called a loading vector) in belief space and λ_j are eigenvalues. We use “loadings” to refer to the entries of $\mathbf{U}$ (or equivalently its columns), while reserving “scores” for the projected individual-level coordinates $t_{ij}^* = \mathbf{u}_j^\top \mathbf{y}_i^*$.

C.1 Rayleigh quotient and projected variance

The variance of the latent belief vector projected onto a unit direction $\mathbf{u}$ is the Rayleigh quotient [Goodman, 2019, Rayleigh, 1896]:

$$\text{Var}(\mathbf{u}^\top \mathbf{y}_i^*) = \mathbf{u}^\top \text{Cov}(\mathbf{y}_i^*) \mathbf{u} = \mathbf{u}^\top \mathbf{R} \mathbf{u}, \quad (3)$$

where $\mathbf{R}$ is the latent correlation matrix. Maximizing $\mathbf{u}^\top \mathbf{R} \mathbf{u}$ over unit vectors yields λ_1 and achieves the maximum at $\mathbf{u}_1$, and similarly for subsequent orthogonal directions.

It is useful to note the equivalence between decomposing $\mathbf{R}$ and decomposing the latent beliefs matrix $\mathbf{Y}^* \in \mathbb{R}^{n \times p}$. We can apply singular value decomposition:

$$\mathbf{Y}^* = \mathbf{W} \mathbf{D} \mathbf{U}^\top;$$

given $\mathbf{R} \propto (\mathbf{Y}^\star)^\top \mathbf{Y}^\star$, we have:

$$(\mathbf{Y}^\star)^\top \mathbf{Y}^\star = \mathbf{U} \mathbf{D}^2 \mathbf{U}^\top,$$

hence the PC loading directions from the eigen-decomposition of $\mathbf{R}$ coincide with the right singular vectors (columns of $\mathbf{U}$) in the SVD of that matrix [Jolliffe and Cadima, 2016].

In our approach we never estimate the latent belief matrix $\mathbf{Y}^\star$, so we do not recover individual-level PC scores. However, the Rayleigh quotient lets us recover the *variance* that any belief system places along any fixed orthonormal basis. This is the key ingredient in our “missing structure” comparison: we treat the GSS principal-component directions (estimated on the persona-conditional belief matrix) as a shared set of human axes, and then evaluate how much variance each LLM belief system exhibits when projected onto those same axes. Reporting these projected variances relative to the GSS benchmark yields Rayleigh ratios, which can be read as dimension-specific *over-* or *under-realization* of human belief structure (Figure 6). Formally, for each GSS direction $\mathbf{u}_j$ we compute $\mathbf{u}_j^\top \mathbf{R}_{\text{LLM}} \mathbf{u}_j$ and divide by the corresponding GSS eigenvalue λ_j .

C.1.1 Constraint Diagnostics

Following from solving the constrained optimisation problem posed by the Rayleigh quotient, λ_j is the variance explained by the j^{th} PC for the belief system encoded by $\mathbf{R}$. The total variance represented by $\mathbf{R}$ is its trace, $\text{tr}(\mathbf{R}) = \sum_{j=1}^p \lambda_j$. For a polychoric correlation matrix, $\text{tr}(\mathbf{R}) = p$ because each latent belief propensity has unit variance. This immediately yields our first scalar diagnostic of constraint, the share of the variance explained by the first principal component:

$$\text{PVE}_1(\mathbf{R}) = \frac{\lambda_1}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_1}{\text{tr}(\mathbf{R})} = \frac{\lambda_1}{p}. \quad (4)$$

Intuitively, $\text{PVE}_1(\mathbf{R})$ is large when the belief-system variance concentrates strongly along a single latent axis, i.e., when many beliefs move together and are mutually predictable through one dominant dimension.

A complementary scalar diagnostic that captures *global* linear dependence is *effective dependence* [Pena and Rodriguez, 2003], defined for a $p \times p$ correlation matrix as:

$$D_e(\mathbf{R}) = 1 - |\mathbf{R}|^{1/p}, \quad (5)$$

where $|\mathbf{R}|$ denotes the determinant. Since determinants are multiplicative and directions $\mathbf{u}$ are orthonormal, we can unify our framework and express $D_e(\mathbf{R})$ as a function of the geometric mean of the eigenvalues:

$$D_e(\mathbf{R}) = 1 - |\mathbf{R}|^{1/p} = 1 - |\mathbf{\Lambda}|^{1/p} = 1 - \left(\prod_{j=1}^p \lambda_j \right)^{1/p}. \quad (6)$$

When beliefs are nearly independent, $\mathbf{R} \approx \mathbf{I}_p$, all $\lambda_j \approx 1$, $|\mathbf{R}| \approx 1$, and $D_e(\mathbf{R}) \approx 0$; stronger dependence produces more uneven eigenvalues, driving some λ_j toward 0 and shrinking $|\mathbf{R}|$, thereby increasing $D_e(\mathbf{R})$ toward 1 [Pena and Rodriguez, 2003].

Both $\text{PVE}_1(\mathbf{R})$ and $D_e(\mathbf{R})$ depend only on the eigenvalues $\{\lambda_j\}$, but they capture different facets of “constraint”: $\text{PVE}_1(\mathbf{R})$ is a *concentration* measure, quantifying how strongly the belief system concentrates variance along its leading eigen-direction, whereas $D_e(\mathbf{R})$ is a *dispersion* measure, quantifying how unevenly the variance in the system is distributed across the orthogonal eigen-directions. Pena and Rodriguez [2003] further show that $D_e(\mathbf{R})$ is closely related to the effective number of independent orthonormal dimensions necessary to explain $\approx 90\%$ of the variance in $\mathbf{R}$. Figure 8 presents the placement of each

belief-system generator along this 2-dimensional plane of constraint. Clearly, systems of beliefs generated persona-based silicon sampling are dramatically over-constrained.

Constraints on Generated Belief Systems

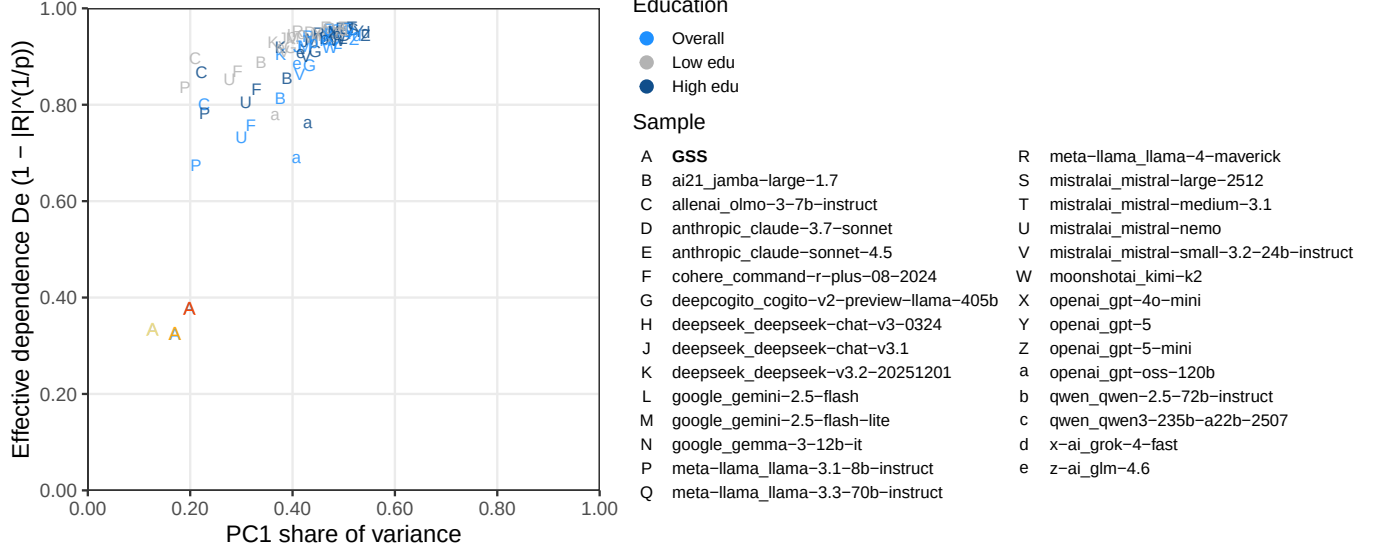


Figure 8: **Placement on the two-dimensional constraint plane.** Each point places a belief-system model (GSS and each LLM) in terms of (i) concentration on a single dominant dimension (x-axis: PVE_1) and (ii) global linear dependence across all items (y-axis: De). Points are shown overall and by education group; the GSS benchmark is highlighted for reference. Movement up and to the right corresponds to increasingly over-constrained belief systems: more variance concentrated on one axis and greater overall dependence among items.

Understanding the differences in these two scalar constraint diagnostic can help us explain the pattern of Education stratification effects in Figure 3. These often appear statistically insignificant and, for some models, weakly inverted relative to the GSS. In the GSS, higher education is associated with *greater* dependence ($\Delta \approx 0.045$), consistent with classic “constraint premium” results; in contrast, most LLMs show Δ s that are near zero and slightly negative. We

caution against over-interpreting this pattern, as many of the LLM Δ s are small in magnitude and often statistically indistinguishable from zero. Nevertheless, one way to reconcile this with Figure 2 is that the two diagnostics emphasize different parts of the eigen-spectrum: PVE_1 primarily tracks concentration on the leading ideological axis, whereas D_e is sensitive to the entire tail of smaller eigenvalues. Substantively, this implies a contrast in how synthetic belief systems are composed across education groups: for lower-educated LLM personas, the weight placed on any single axis is typically lower, but a somewhat smaller number of axes suffices to describe most of the system—belief variance is spread more evenly across fewer dimensions. For higher-educated personas, by contrast, beliefs are more strongly organized around one dominant ideological axis (raising PVE_1), yet the remaining variance is distributed across a somewhat broader set of weaker dimensions. We should stress again that these differences in the education gradient of the D_e constraint are generally very small.

D Persona Mediation Analysis

Persona-based prompting conditions responses on demographic traits, which can act as strong covariates linking multiple attitudes. If a belief system is heavily *persona-mediated*, then much of its apparent inter-attitude dependence should be explained by persona variables, and the residual association structure among beliefs should weaken substantially once those persona variables are partialled out.

We formalize this using a block partition of the full latent correlation matrix,

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{\mathcal{B}\mathcal{B}} & \mathbf{R}_{\mathcal{B}\mathcal{P}} \\ \mathbf{R}_{\mathcal{P}\mathcal{B}} & \mathbf{R}_{\mathcal{P}\mathcal{P}} \end{bmatrix}, \quad \mathbf{R}_{\mathcal{B}\mathcal{B}} \in \mathbb{R}^{o \times o}, \mathbf{R}_{\mathcal{P}\mathcal{P}} \in \mathbb{R}^{q \times q}, \quad (7)$$

where $\mathcal{B}$ indexes the o beliefs and $\mathcal{P}$ indexes the q persona variables. Under the joint latent-Gaussian model for $(\mathbf{y}_i^{\star\mathcal{B}}, \mathbf{y}_i^{\star\mathcal{P}})$, the conditional covariance of latent beliefs given persona is given by the Schur complement:

$$\boldsymbol{\Sigma}_{\mathcal{B}|\mathcal{P}} := \text{Cov}(\mathbf{y}_i^{\star\mathcal{B}} | \mathbf{y}_i^{\star\mathcal{P}}) = \mathbf{R}_{\mathcal{B}\mathcal{B}} - \mathbf{R}_{\mathcal{B}\mathcal{P}} \mathbf{R}_{\mathcal{P}\mathcal{P}}^{-1} \mathbf{R}_{\mathcal{P}\mathcal{B}}. \quad (8)$$

Because (8) is a covariance (not a correlation) matrix, we convert it to a residual (partial) correlation matrix by standardizing:

$$\mathbf{R}_{\mathcal{B}|\mathcal{P}} := \mathbf{D}^{-1/2} \boldsymbol{\Sigma}_{\mathcal{B}|\mathcal{P}} \mathbf{D}^{-1/2}, \quad \mathbf{D} = \text{diag}(\boldsymbol{\Sigma}_{\mathcal{B}|\mathcal{P}}). \quad (9)$$

The matrix $\mathbf{R}_{\mathcal{B}|\mathcal{P}}$ captures the residual interlocking structure among beliefs *after* accounting for persona-level covariance; equivalently, its off-diagonal entries are partial correlations between pairs of beliefs controlling for all persona variables.

Influence over Constraint Diagnostics. Having obtained $\mathbf{R}_{\mathcal{B}|\mathcal{P}}$, we can quantify (i) *residual constraint* among beliefs via the same eigenvalue-based diagnostics, e.g. $\text{PVE}_1(\mathbf{R}_{\mathcal{B}|\mathcal{P}}) = \lambda_1(\mathbf{R}_{\mathcal{B}|\mathcal{P}})/o$ and $D_e(\mathbf{R}_{\mathcal{B}|\mathcal{P}}) = 1 - |\mathbf{R}_{\mathcal{B}|\mathcal{P}}|^{1/o}$, and (ii) *persona-mediated constraint* as the reduction in dependence induced by conditioning:

$$\Delta D_e := D_e(\mathbf{R}_{\mathcal{B}\mathcal{B}}) - D_e(\mathbf{R}_{\mathcal{B}|\mathcal{P}}), \quad \Delta \text{PVE}_1 := \text{PVE}_1(\mathbf{R}_{\mathcal{B}\mathcal{B}}) - \text{PVE}_1(\mathbf{R}_{\mathcal{B}|\mathcal{P}}). \quad (10)$$

Large values of $(\Delta D_e, \Delta \text{PVE}_1)$ indicate that a substantial fraction of apparent belief constraint is carried by persona-level covariation (i.e., beliefs become markedly less mutually predictable once persona is accounted for), whereas small values indicate that beliefs remain tightly interlocked even after conditioning,

consistent with a more inherently interlocked belief system.

Figure 9 contrasts belief-system constraint measured on the unconditional belief block, $\mathbf{R}_{\mathcal{B}\mathcal{B}}$, with the corresponding persona-conditional (residual) belief structure, $\mathbf{R}_{\mathcal{B}|\mathcal{P}}$. The gap between these two quantities isolates how much of the apparent belief structure is carried by socio-demographic covariation (see also Figure 4). Changes in D_e are smaller and more heterogeneous ($\Delta D_e \approx -0.05$ to 0.14) than we see in Figure 4: most models exhibit a modest decrease in global dependence after conditioning, while a small number of models show little change or slight increases, suggesting that persona variables play a weaker and less consistent role in shaping the dispersion of beliefs across multiple axes than they do in shaping the dominant axis.

Decomposing Belief-system Variance. Beyond paired scalar diagnostics, we can decompose variance in the belief block $\mathbf{R}_{\mathcal{B}\mathcal{B}}$ into the component mediated by persona and the residual component among beliefs. From the Schur complement in Eq. (8), we obtain the additive decomposition:

$$\mathbf{R}_{\mathcal{B}\mathcal{B}} = \underbrace{\mathbf{R}_{\mathcal{B}\mathcal{P}}(\mathbf{R}_{\mathcal{P}\mathcal{P}})^{-1}\mathbf{R}_{\mathcal{P}\mathcal{B}}}_{\Sigma_{\text{persona}}} + \Sigma_{\mathcal{B}|\mathcal{P}}. \quad (11)$$

Let $\mathbf{R}_{\mathcal{B}\mathcal{B}} = \mathbf{U}_{\mathcal{B}}\mathbf{\Lambda}_{\mathcal{B}}\mathbf{U}_{\mathcal{B}}^{\top}$ with eigenpairs $(\lambda_{\mathcal{B},j}, \mathbf{u}_{\mathcal{B},j})$, where $\mathbf{\Lambda}_{\mathcal{B}} = \text{diag}(\lambda_{\mathcal{B},1}, \dots, \lambda_{\mathcal{B},o})$. The total variance explained along belief PC j is $\lambda_{\mathcal{B},j} = \mathbf{u}_{\mathcal{B},j}^{\top} \mathbf{R}_{\mathcal{B}\mathcal{B}} \mathbf{u}_{\mathcal{B},j}$. We attribute to persona variables the following Rayleigh quotient

$$v_{\text{persona},j} := \mathbf{u}_{\mathcal{B},j}^{\top} \Sigma_{\text{persona}} \mathbf{u}_{\mathcal{B},j}; \quad (12)$$

Effective dependence among beliefs: unconditional vs conditional on persona

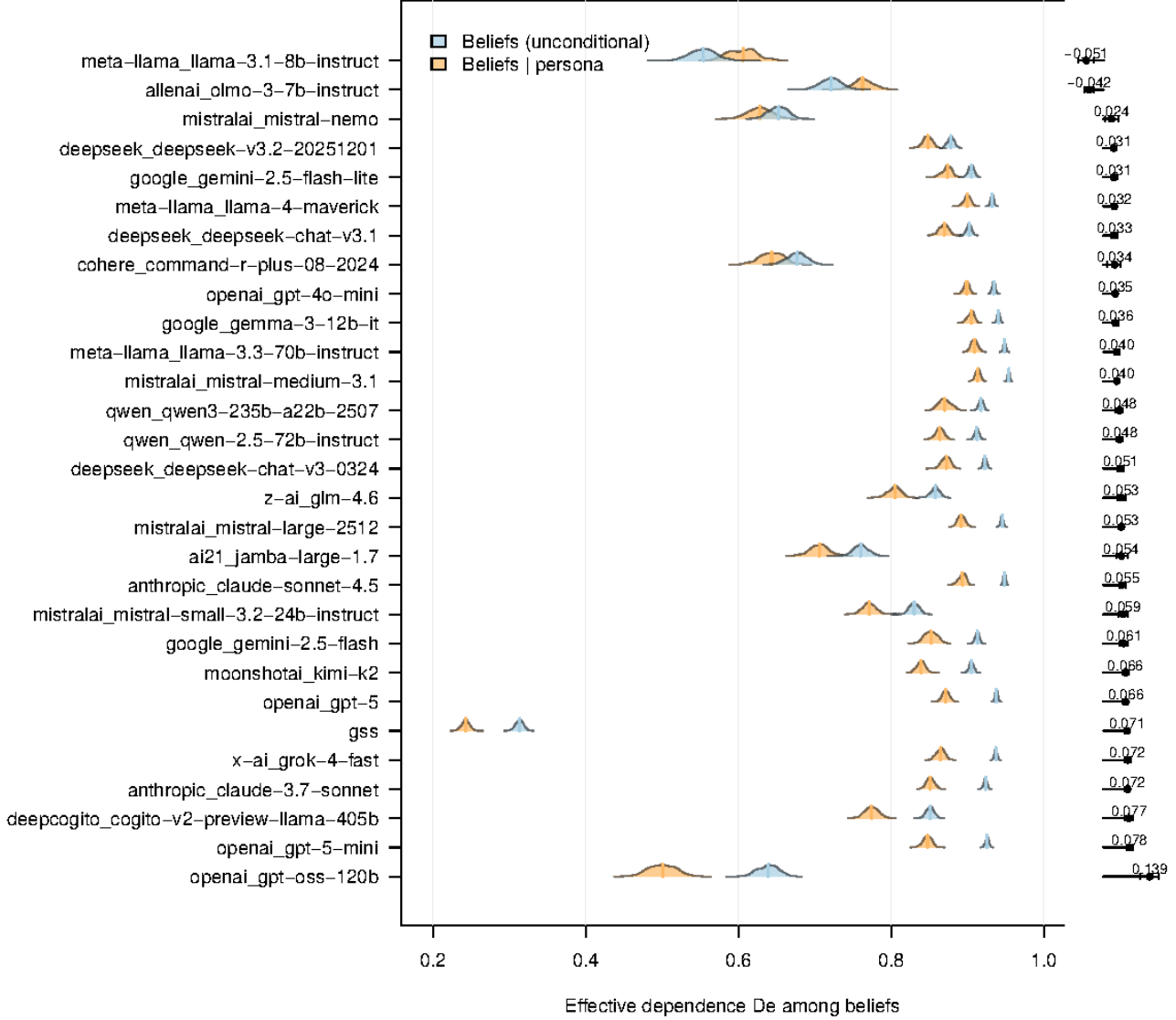


Figure 9: **Persona mediation of global belief dependence.** Bootstrap distributions of effective dependence among beliefs, D_e , computed on the unconditional belief block $\mathbf{R}_{BB}$ and on the persona-conditional residual belief matrix $\mathbf{R}_{B|P}$. Right-side lollipops show $\Delta D_e = D_e(\mathbf{R}_{BB}) - D_e(\mathbf{R}_{B|P})$: the point is the median and whiskers denote the 95% bootstrap interval. Lollipops are dark when the 95% interval excludes zero (light otherwise). Positive values indicate that conditioning on persona reduces global linear dependence among beliefs; negative values indicate slightly higher residual dependence after conditioning.

and define the residual (non-persona) contribution along the same direction as:

$$v_{\text{resid},j} := \mathbf{u}_{\mathcal{B},j}^\top \boldsymbol{\Sigma}_{\mathcal{B}|\mathcal{P}} \mathbf{u}_{\mathcal{B},j} = \lambda_{\mathcal{B},j} - v_{\text{persona},j}. \quad (13)$$

Normalising by total belief variance $\text{tr}(\mathbf{R}_{\mathcal{B}\mathcal{B}}) = o$, we define cumulative persona and residual shares of explained belief variance up to PC m as

$$S_{\text{persona}}(m) := \frac{1}{o} \sum_{j=1}^m v_{\text{persona},j}, \quad S_{\text{resid}}(m) := \frac{1}{o} \sum_{j=1}^m v_{\text{resid},j}, \quad (14)$$

so that $S_{\text{persona}}(m) + S_{\text{resid}}(m) = \frac{1}{o} \sum_{j=1}^m \lambda_{\mathcal{B},j}$. The cumulative curves enable a holistic comparison of the composition of the belief system. A level difference between the persona and residual shares of explained variance appears as a persistent gap between the two curves, indicating whether persona-level dependence dominates the system or is subordinate to deeper interlocking structure among beliefs. Differences in slope capture how quickly each component accumulates explained variance (i.e., how rapidly the curves converge to their final cumulative %): steep early gains imply concentration in the leading principal components, whereas gradual accumulation implies dispersion across many components. Together, these features identify not only the overall degree of persona mediation, but also where (in the principal-component spectrum) persona-related versus residual belief variance is concentrated.

Figure 5 presents the decomposition of the cumulative share of explained variance by persona and non-persona covariates. Figure 10 provides a complementary view of belief-system constraint using the eigen-spectrum of the unconditional belief correlation matrix. For each generator, we plot the cumulative proportion of belief variance explained as additional principal components are included. Systems whose variance concentrates quickly into the first few compo-

nents exhibit a more low-dimensional (and hence more “constrained”) geometry, whereas slower accumulation indicates more diffuse, higher-dimensional structure. The vertical reference markers summarize this intuition by indicating the smallest number of components required to reach 90% cumulative explained variance.

Relative to the GSS benchmark, many LLM persona systems reach 90% variance with fewer components, consistent with the idea that synthetic belief profiles collapse onto a smaller set of latent dimensions. Stratification by education shows the expected pattern in the human data—higher-education respondents tend to exhibit a more compact structure (fewer components needed for a given variance threshold)—but this gradient is often attenuated or irregular across LLMs.

Identifying *Missing Structure*.

The cumulative explained-variance curves tell us *how much* structure differs across belief-system generators, but not *what kind* of structure is absent or over-emphasised. To identify *which* substantive dimensions are “missing” in LLM-generated belief systems, we compare generators along a *shared* orthonormal basis derived from the GSS itself. We build a GSS principal-component basis on the persona-conditional polychoric correlation matrix to obtain loading directions $\{\mathbf{u}_j\}$ and eigenvalues $\{\lambda_j\}$. This provides a human benchmark basis: each $\mathbf{u}_j$ is a coherent axis of covariation among beliefs in the GSS, and λ_j quantifies how much GSS variance lies along that axis. We then compute the variance that each LLM-generated belief system exhibits along each GSS axis via the Rayleigh quotient

$$v_{LLM,j} = \mathbf{u}_j^\top \mathbf{R}_{LLM} \mathbf{u}_j,$$

where $\mathbf{R}_{LLM}$ is the LLM’s persona-conditional correlation matrix. We summa-

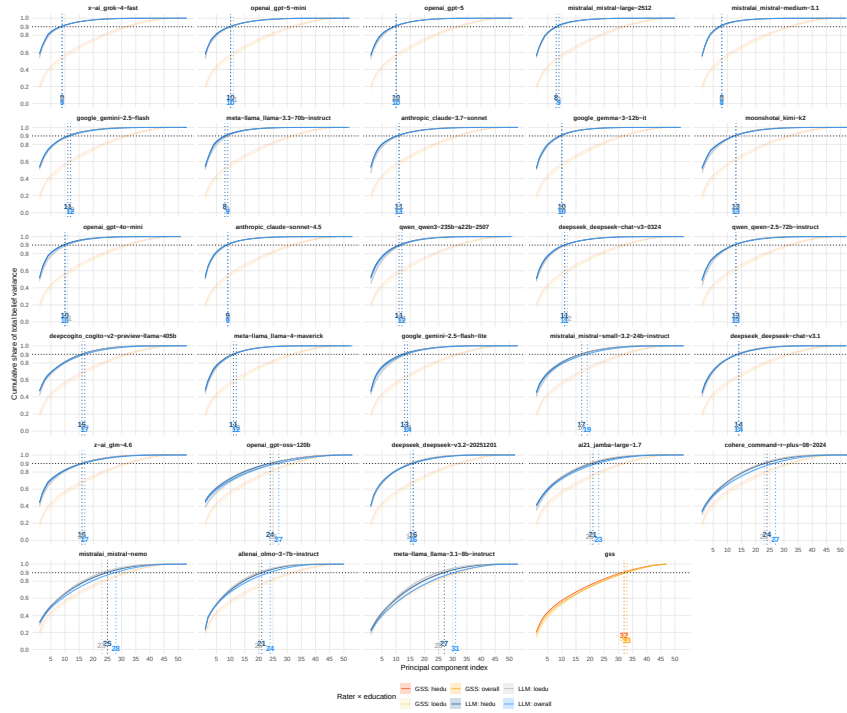


Figure 10: **Cumulative explained variance in belief systems.** For each belief-system generator, we present the cumulative share of total belief variance explained by successive principal components of the unconditional belief correlation matrix. Curves are shown overall and stratified by education, separately for the human benchmark (GSS) and for LLM personas. Vertical dotted lines and bold numeric labels indicate the minimum number of components required to reach 90% cumulative explained variance.

size *relative* variance using the Rayleigh ratio

$$\text{RR}_{LLM,j} = \frac{v_{LLM,j}}{\lambda_j},$$

which can be interpreted as “how much variance a given LLM-generated belief system has on GSS component j relative to humans.” Values $\text{RR}_{LLM,j} \approx 1$ indicate that the rater matches the human benchmark on that dimension; $\text{RR}_{LLM,j} < 1$ indicates that the rater *under-realizes* (attenuates) that human dimension; and $\text{RR}_{LLM,j} > 1$ indicates *over-realization* (amplification).

This comparison operationalises *missing structure* as dimension-specific variance deficits relative to the GSS (Figure 6). Substantively, to determine *what* is missing, we inspect the loading vectors $\mathbf{u}_j$ for PCs where LLMs systematically show $\text{RR}_{LLM,j} \ll 1$. When LLMs under-express variance in higher-order components that exist in the GSS (i.e., components beyond the dominant ideological axis), this indicates that LLM-generated belief-systems lack the weaker, cross-cutting, or domain-specific bundles of association that characterize real publics – secondary cleavages that link subsets of issues without collapsing them into a single left–right line.

Figure 11 provides the substantive interpretation for the GSS principal directions used in the Rayleigh-ratio “missing structure” analysis. Each column is a GSS principal component computed from the persona-conditional belief correlation matrix, and each row is an attitude item. Cell shading encodes the absolute loading magnitude $|u_{kj}|$, while the printed values retain the signed loadings; because PC signs are arbitrary, interpretation should focus on the bundles of items with large magnitude and their relative polarity within a column rather than on sign alone.

The figure is most useful as a lookup table: when a model under-realizes a given human component in Figure 6 (Rayleigh ratios < 1), the corresponding

column here indicates which beliefs define that dimension in the human benchmark. Components with loadings concentrated on a narrow subset of items typically reflect domain-specific or cross-cutting issue bundles (rather than the dominant general axis). When LLMs attenuate these higher-order PCs, they effectively compress that finer-grained multidimensional structure into the leading dimension, yielding belief systems that are simultaneously more one-dimensional and less faithful to the secondary organization present in the GSS.

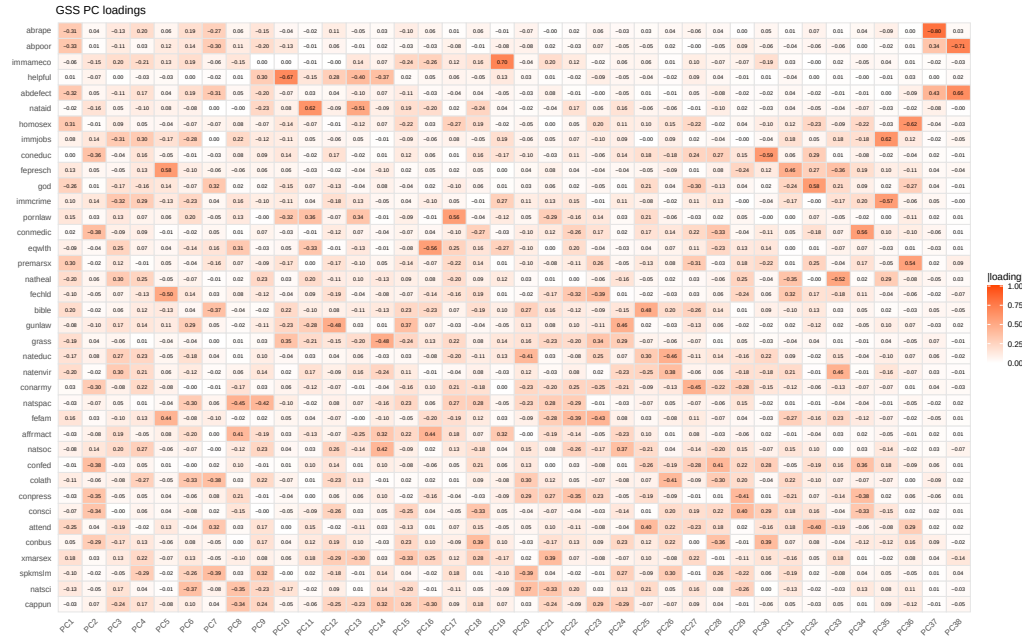


Figure 11: **GSS principal-component loadings used to interpret missing dimensions.** Heatmap of GSS principal-component loadings for the same persona-conditional belief system basis used in Figure 6. Rows correspond to attitude items and columns to GSS PCs; cell shading encodes the absolute loading magnitude $|u_{kj}|$ (darker = larger contribution of item k to component j), while the numeric annotations report signed loadings. *Missing* higher order dimensions in LLMs from Figure 6 can be interpreted by reading across the corresponding column here: concentrated high-magnitude loadings indicate narrower, domain-specific bundles of beliefs that are present in the human benchmark but attenuated in synthetic personas.

E Effect of Constraint on Belief-System Realism

To determine whether higher constraint translates into more (or less) human-like response patterns, we treat each `(persona_id, variable)` pair as a comparable case, and each belief-system generator (an LLM or the GSS) as a “rater” that assigns an ordinal response to that case. To quantify similarity in response profiles, we compute quadratically (squared) weighted Cohen’s κ , which is appropriate for ordered response categories [Gamer et al., 2019]. For each pair of generators, we compute κ on the subset of cases where both provide non-missing responses. We report point estimates only: the case base is extremely large (personas $\times$ items), so uncertainty on each pairwise κ is negligible for our purposes.

Figure 12 reports pairwise agreement across all generators. Overall, LLMs tend to agree positively with the human benchmark, but the LLM–GSS agreement is systematically weaker than LLM–LLM agreement, and remains below commonly used agreement heuristics (often taken to begin around $\kappa \approx 0.8$ and above) [McHugh, 2012]. A small number of lower-capacity models also show weak agreement with other LLMs and extremely low agreement with the GSS, suggesting that they fail to reproduce even the within-family regularities present in larger models.

To relate realism to constraint more directly, we combine Figure 12 with Figure 13 and regress each model’s GSS agreement, $\kappa_{\text{LLM}, \text{GSS}}$, on our constraint diagnostics. With the caveat that the sample of LLMs is modest (though close to exhaustive for the set of models we evaluate), the pattern is notably positive: more constrained models tend to exhibit higher agreement with the human benchmark. Though seemingly counterintuitive, this is consistent with a training dynamic in which concentrating variance onto a few dominant dimensions can reduce loss because a relatively small number of axes explains much of the

Pairwise agreement: Cohen's kappa (weight=squared)

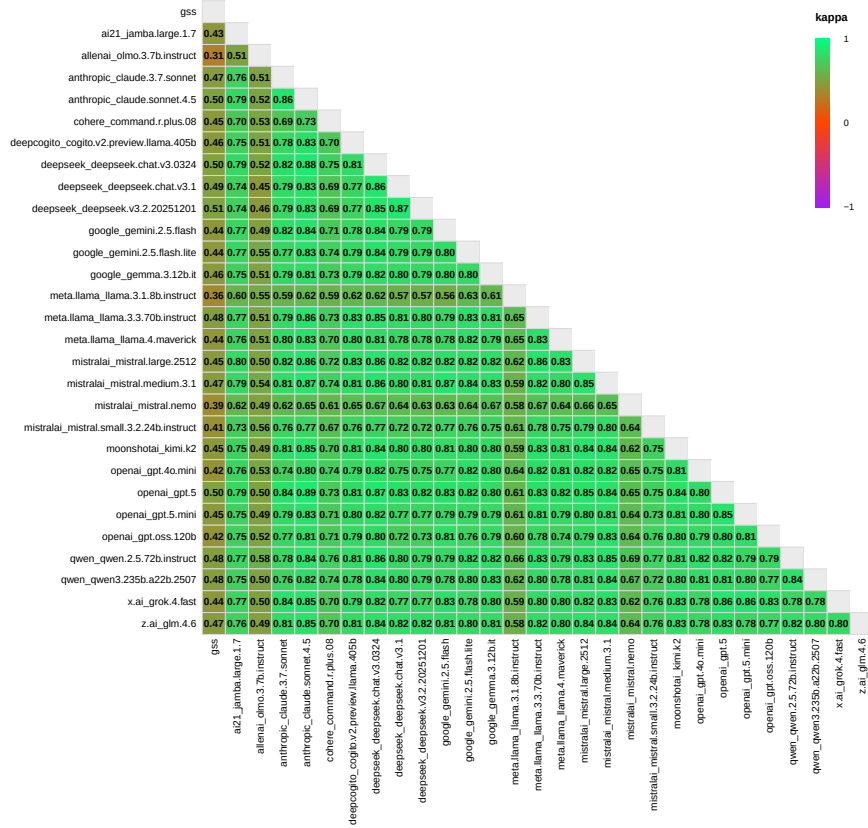


Figure 12: **Pairwise agreement on belief items.** Heatmap of squared-weighted Cohen’s κ computed on belief pairs across personae and items.

GSS variance. Matching the weaker, higher-order dimensions of human belief structure is harder, and theory/empirics in modern deep learning suggest that performance can exhibit *double descent* patterns, rather than improving monotonically [Belkin et al., 2019]. In this understanding, the positive association we observe may reflect that current models are still in a regime where increased constraint tracks improved fit to the dominant human dimensions, while faithful recovery of higher-order structure remains currently out of reach.

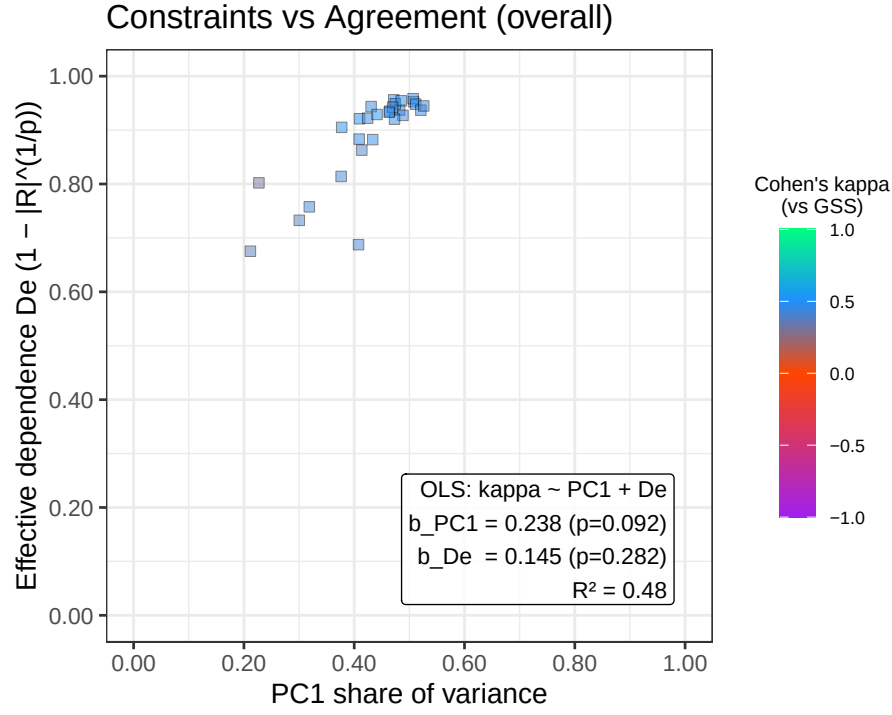


Figure 13: **Constraints versus agreement with the GSS (overall; point estimates).** Scatter of LLM generators in the two-dimensional constraint plane (x-axis: PVE_1 ; y-axis: D_e), using overall-group constraint point estimates. Each square is filled by the model's squared-weighted Cohen's κ against the GSS extracted from the pairwise agreement matrix ($\kappa_{\text{LLM}, \text{GSS}}$). Panel annotation reports the OLS fit coefficients, p-values, and R^2 .

F Missingness stability checks

Because the GSS battery contains substantial item-level missingness (primarily due to survey design and skip patterns), we assess whether our bootstrap + MICE procedure materially alters the estimated association structure. Figure 14 reports two complementary diagnostics.

In the left panels, each point corresponds to a variable pair (j, ℓ) and plots the “raw” polychoric correlation estimated using available cases (pairwise complete observations) against the corresponding correlation estimated after bootstrap resampling and MICE completion (summarized by the average across imputations/bootstraps). Points lying close to the 45-degree line indicate that imputation preserves the baseline correlation geometry, while systematic departures would indicate shrinkage, inflation, or sign reversals induced by the imputation model.

In the right panels, we plot the absolute deviation between the raw and bootstrap+MICE correlations, $\left| \hat{\rho}_{j\ell}^{\text{raw}} - \hat{\rho}_{j\ell}^{\text{MICE}} \right|$, against the fraction of observations missing for that variable pair. This makes explicit how imputation sensitivity scales with information loss: larger discrepancies are expected when pairwise overlap is low, whereas stability under moderate missingness suggests that the imputation model is not driving the substantive structure.

Across replications, differences are generally small for pairs with substantial overlap and increase gradually as pairwise missingness approaches its upper range, consistent with the imputation acting primarily as a variance-stabilizing completion mechanism rather than imposing a new dependence structure.

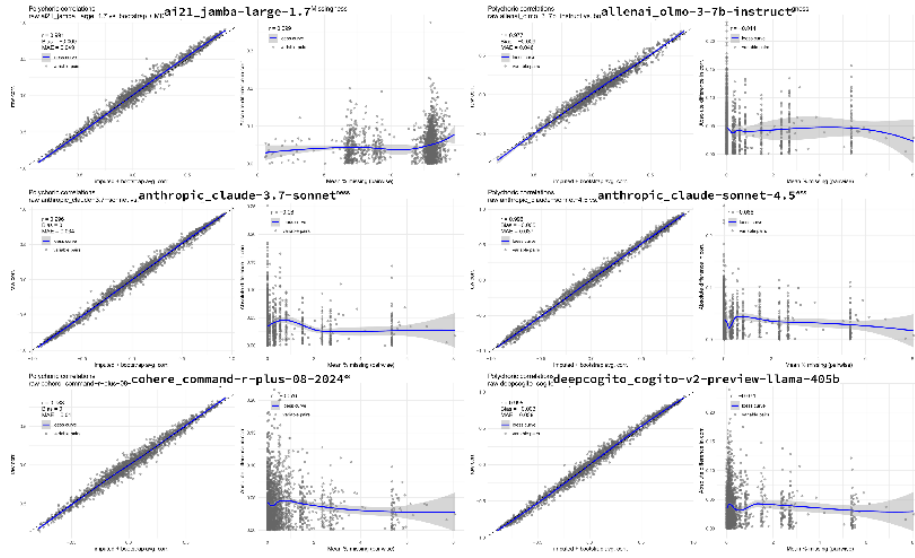


Figure 14: MICE imputation stability checks. Left panels compare raw polychoric correlations against bootstrap + MICE averaged correlations. Right panels show absolute difference vs pairwise missingness. Each row represents one model.

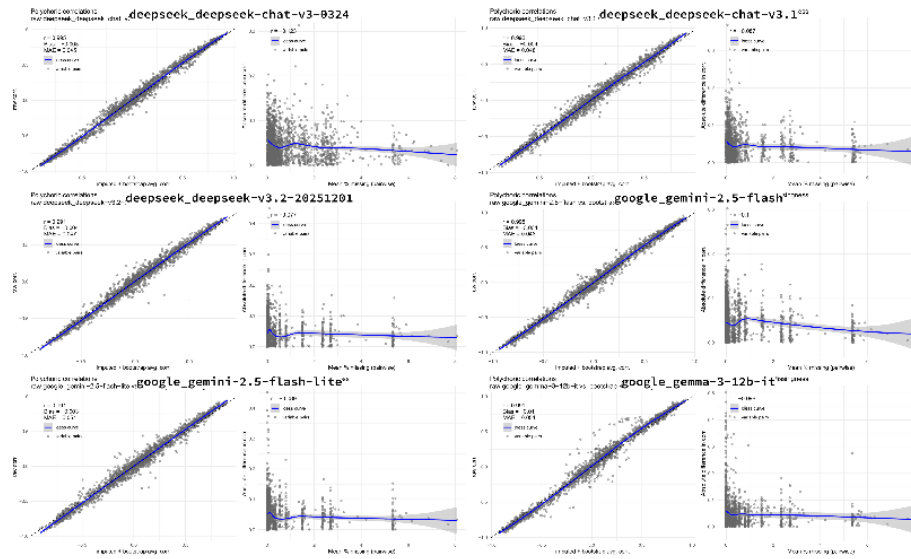


Figure 15: (continued, part 2/5)

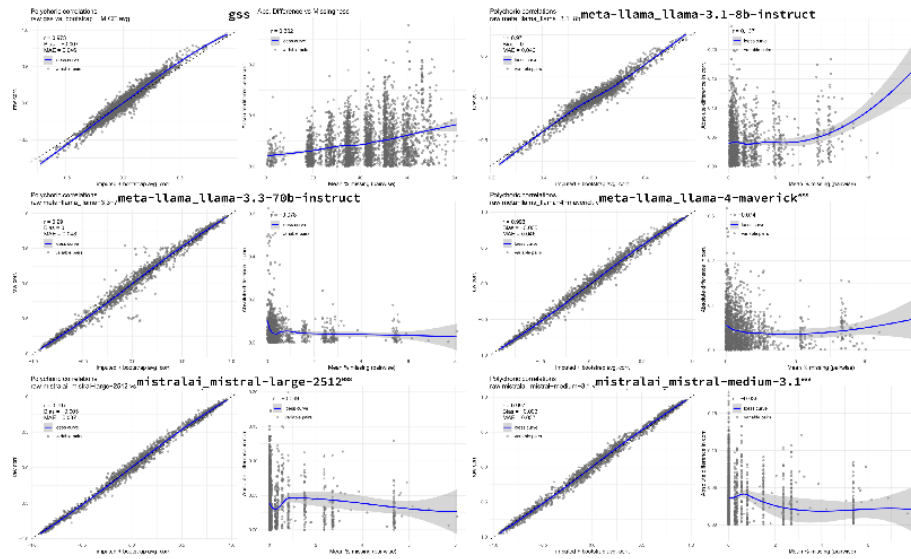


Figure 16: (continued, part 3/5)

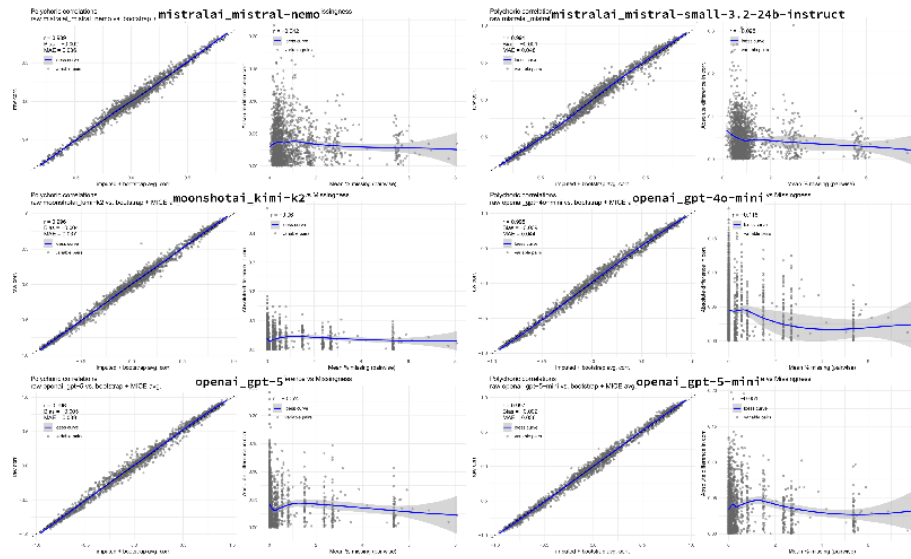


Figure 17: (continued, part 4/5)

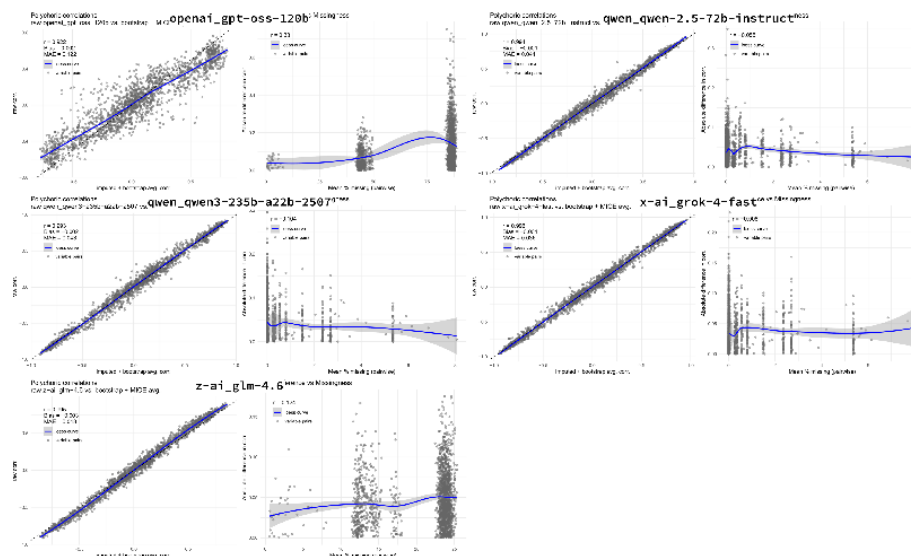


Figure 18: (continued, part 5/5)

G Visualising System Constraints: Complete *Saturn* Plots

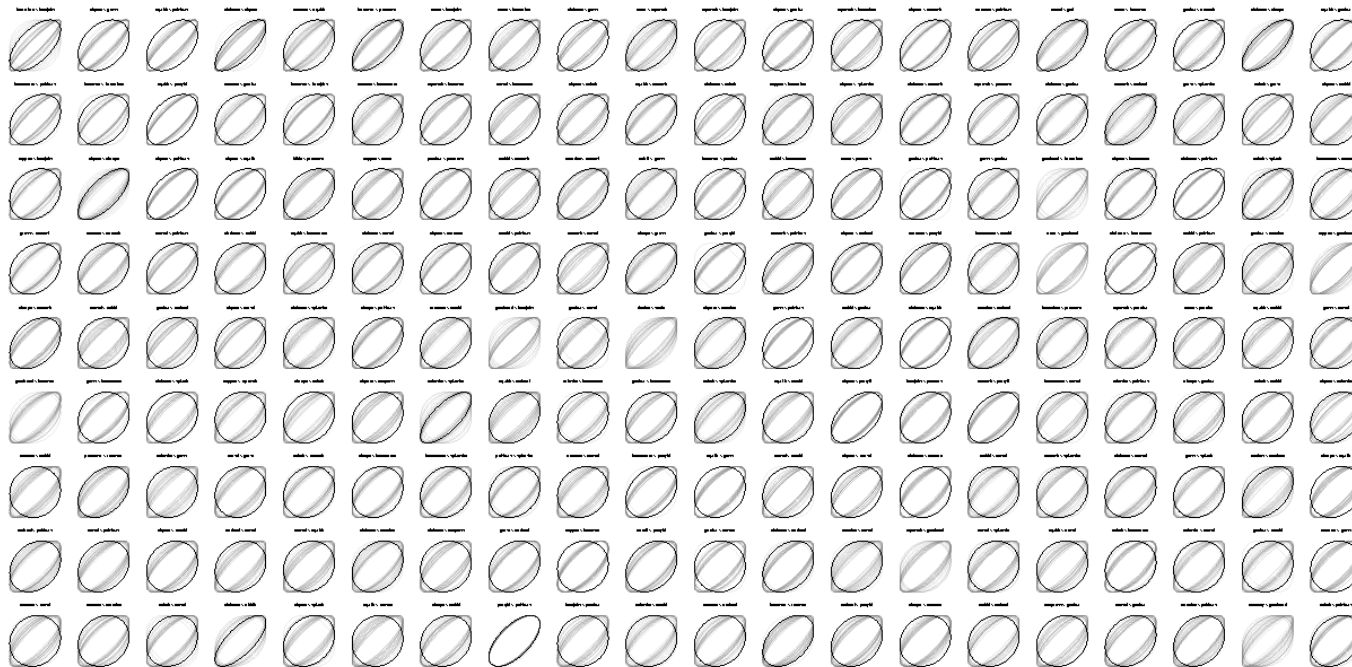


Figure 19: Pairwise belief constraint contours. Each panel shows one variable pair with LLM responses (grey) and GSS human baseline (black). Ellipses represent 50% probability mass of a bivariate normal with correlation set to the median bootstrap polychoric correlation. Ordered from strongest positive to strongest negative correlation.

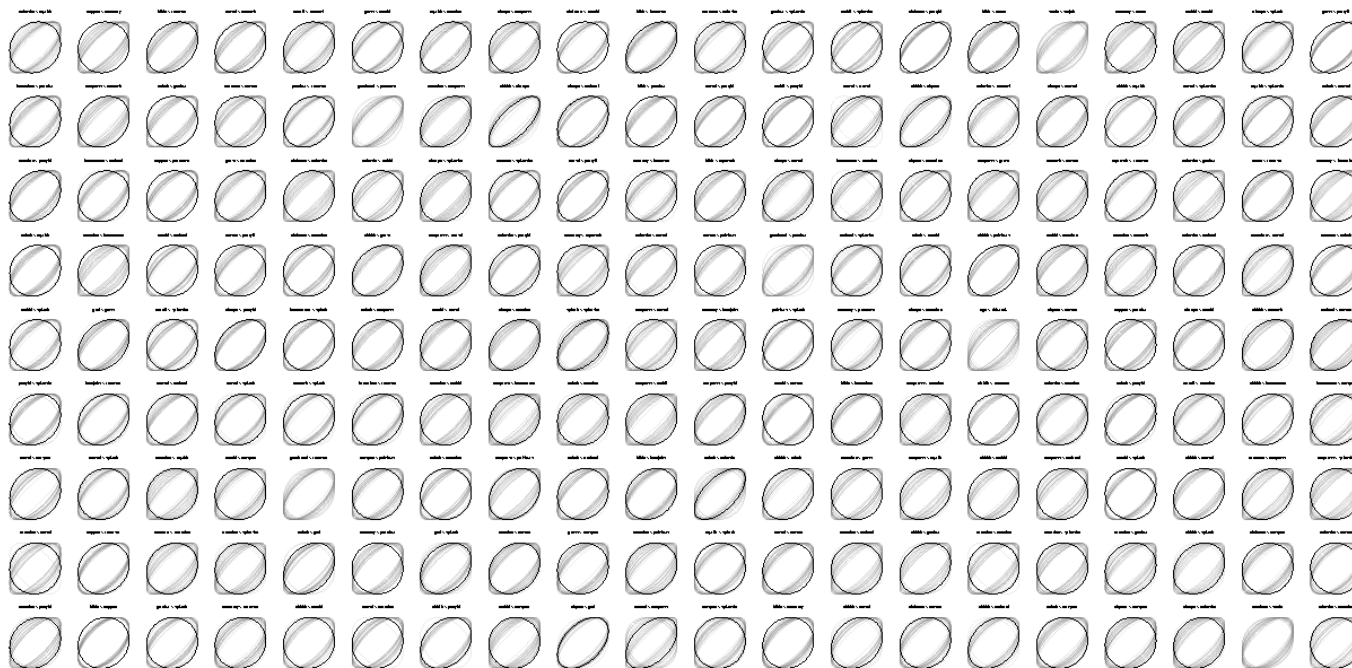


Figure 20: (continued, part 2/11)

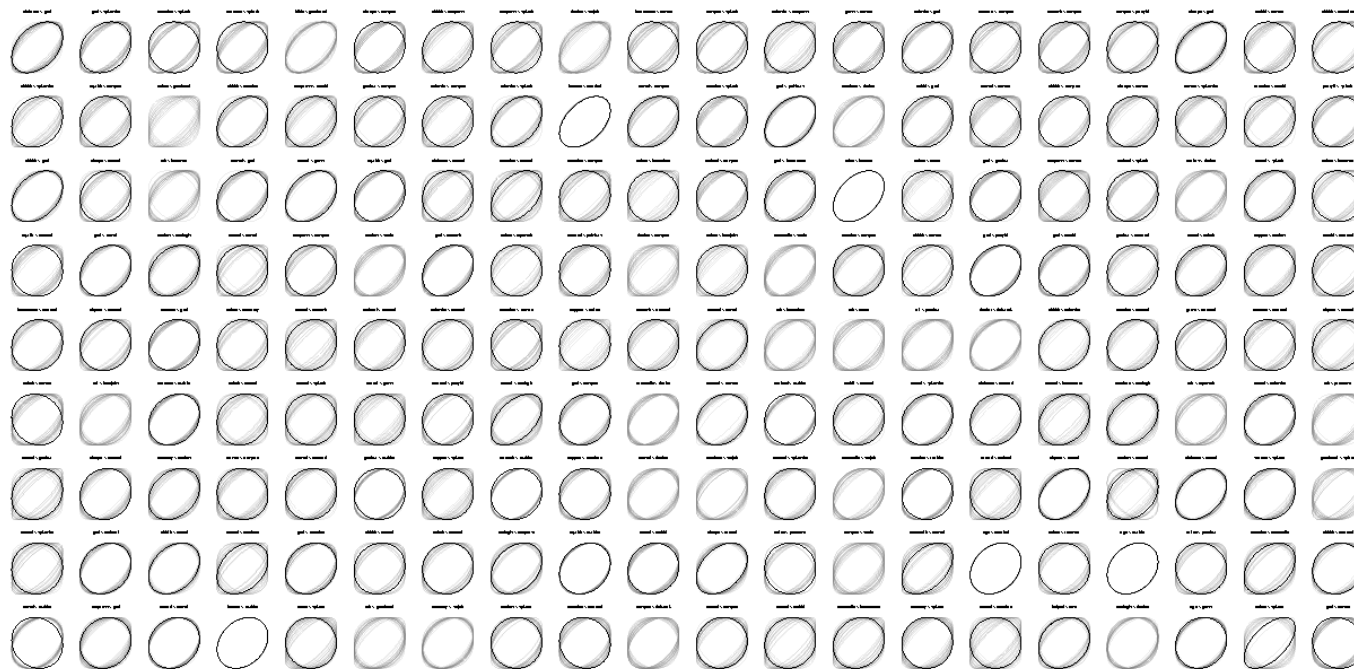


Figure 21: (continued, part 3/11)

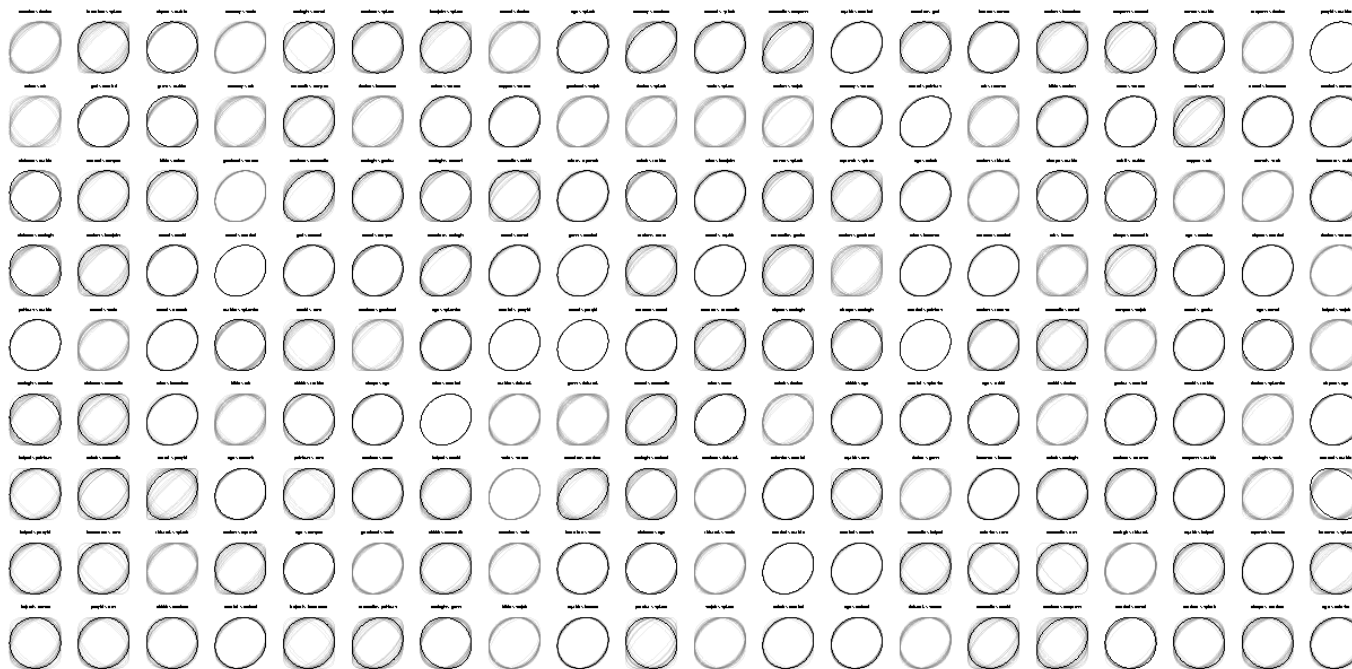


Figure 22: (continued, part 4/11)

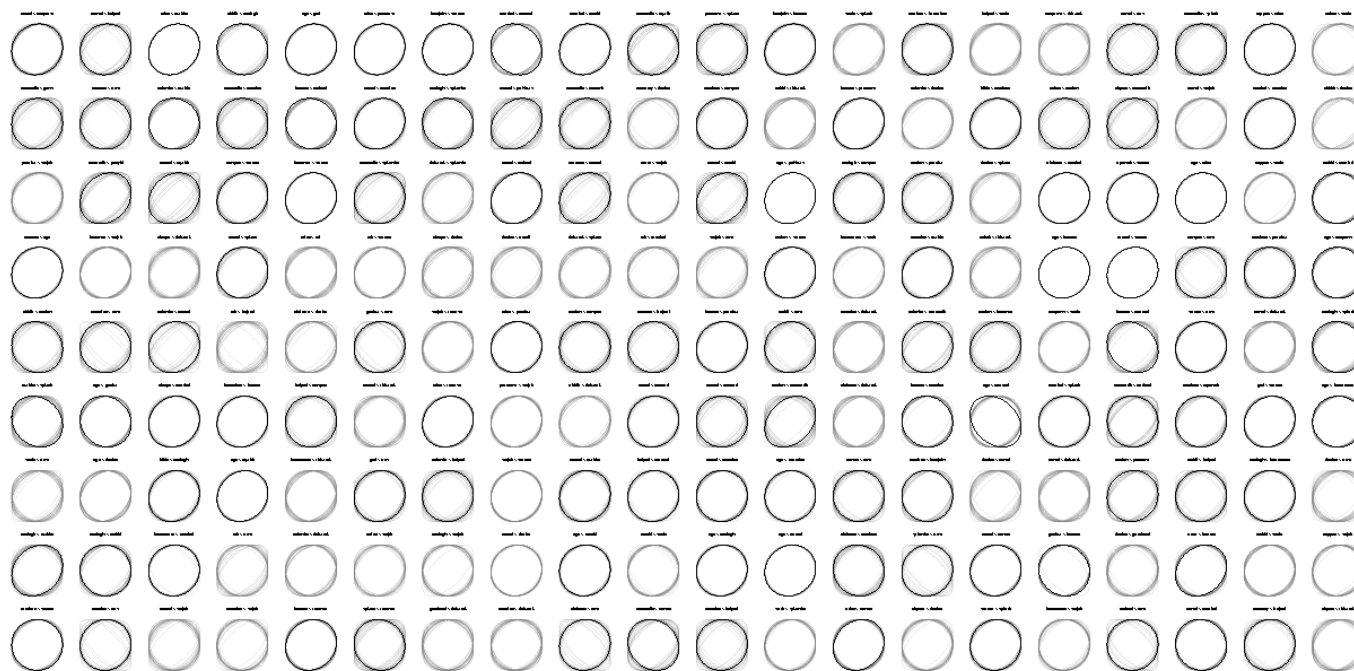


Figure 23: (continued, part 5/11)

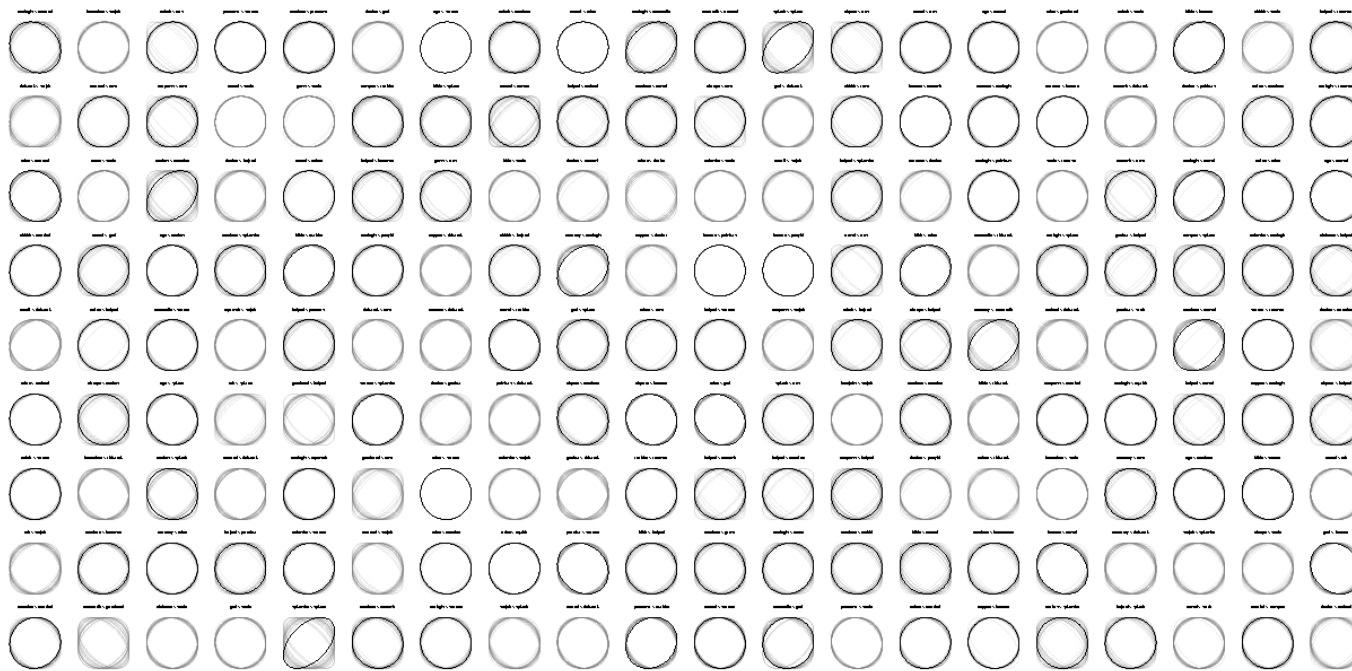


Figure 24: (continued, part 6/11)

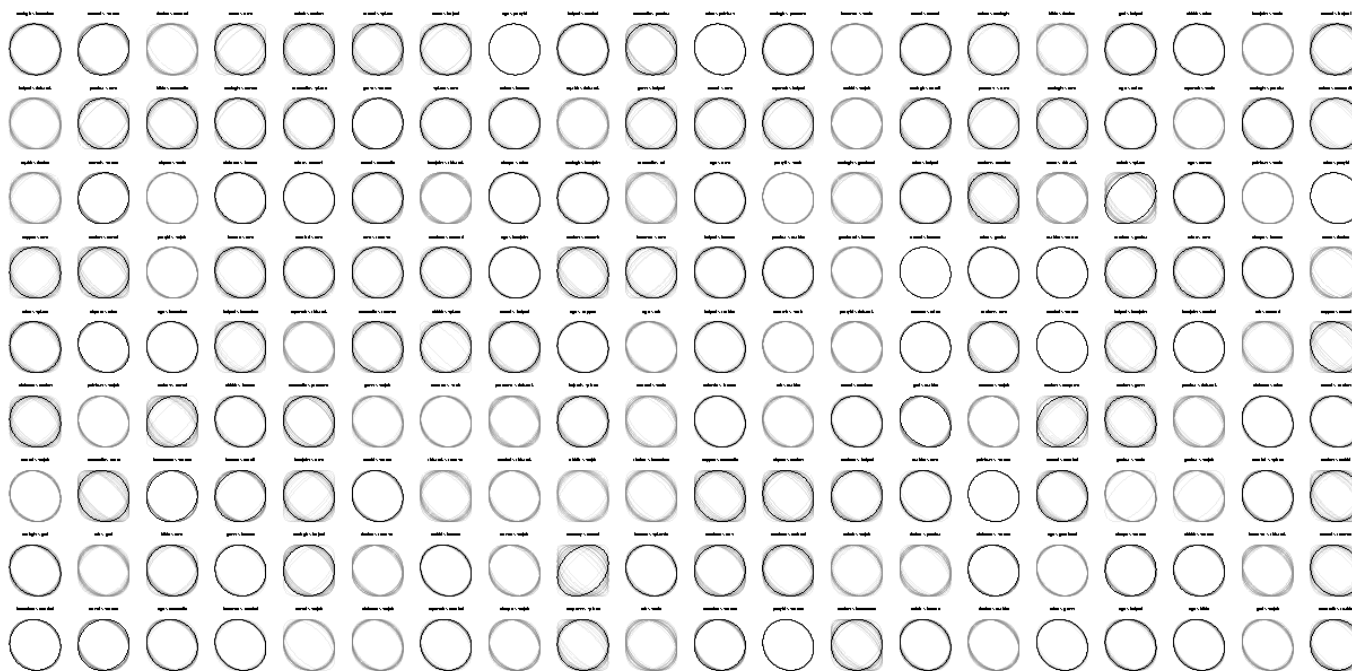


Figure 25: (continued, part 7/11)

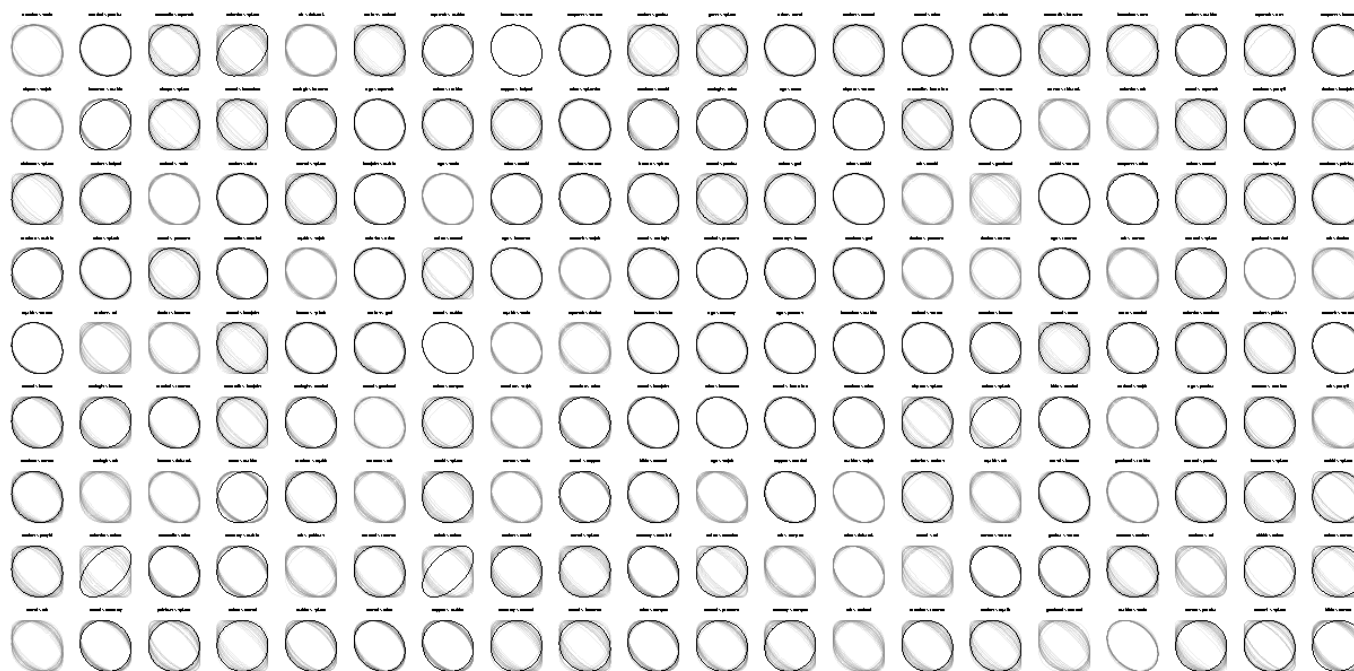


Figure 26: (continued, part 8/11)

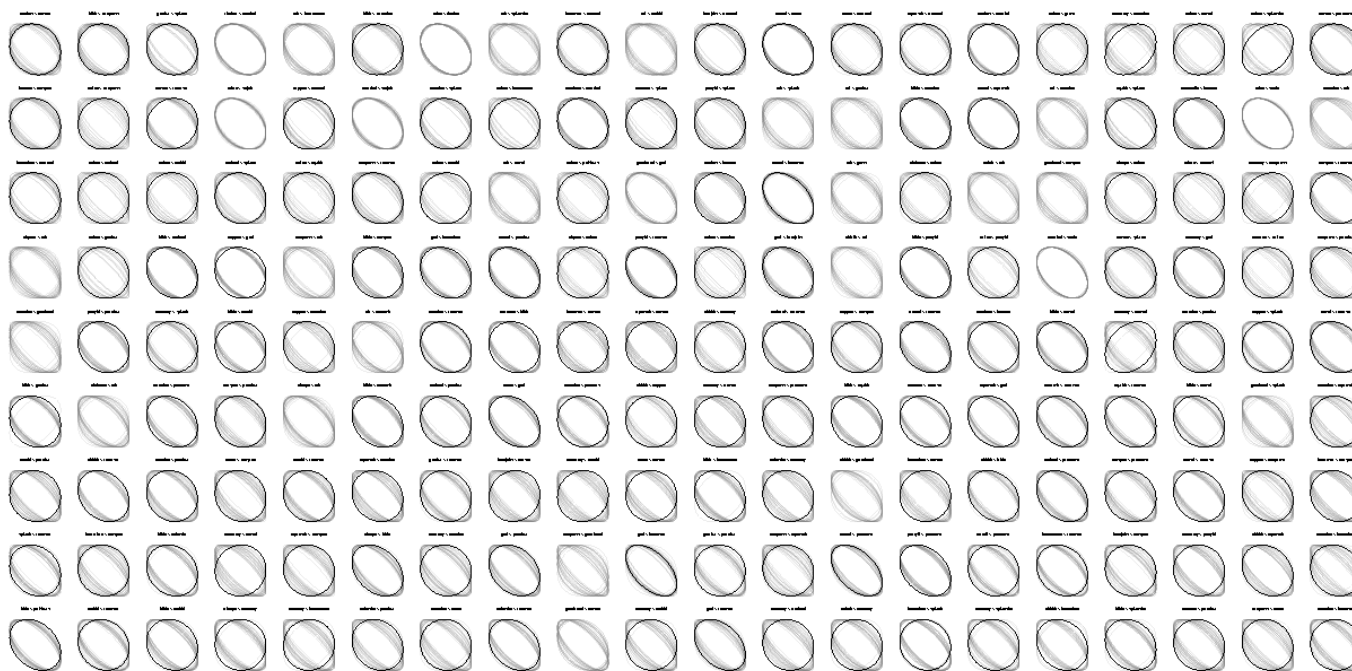


Figure 27: (continued, part 9/11)

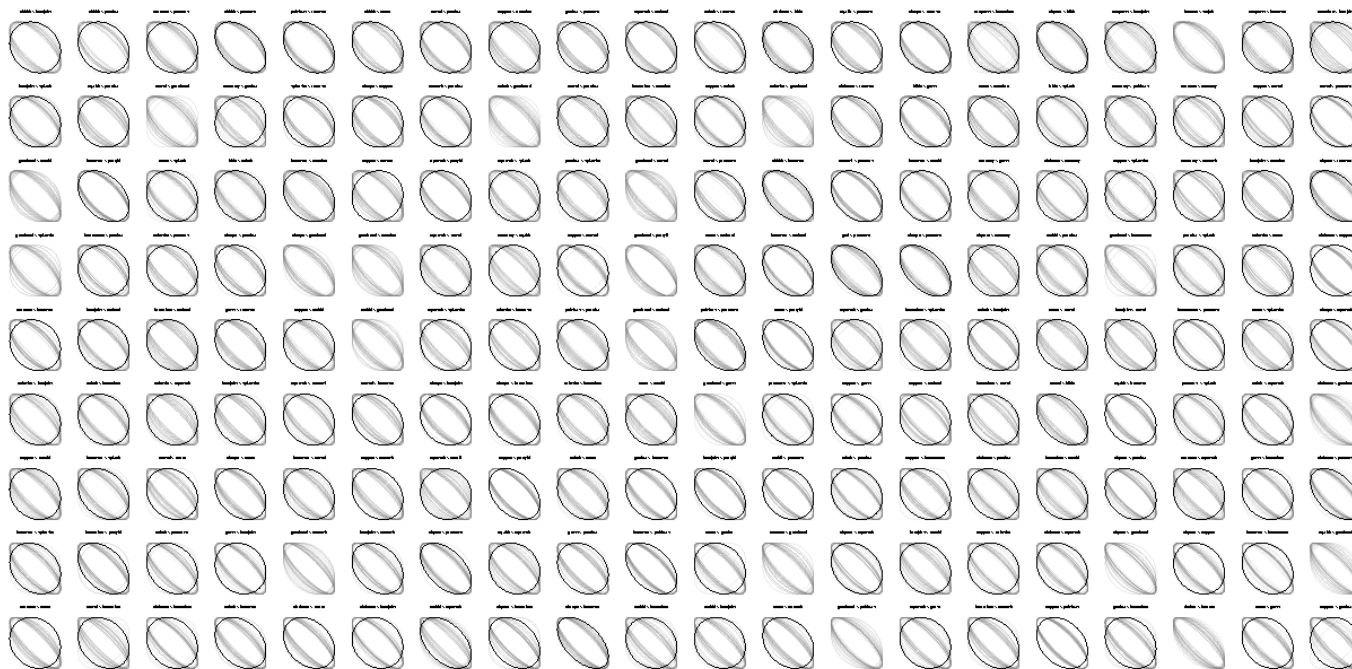


Figure 28: (continued, part 10/11)

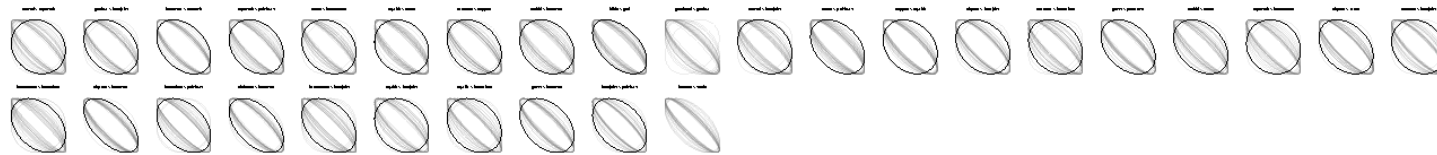


Figure 29: (continued, part 11/11)

H Appendix: GSS Survey Questions

This appendix lists all 52 General Social Survey (GSS) questions administered to LLMs. Questions are organized into two batteries: culture-war items (30 questions) and non-culture-war items (22 questions).

Part 1: Culture-War Items (30 questions)

Abortion (4 items)

1. **abhlth**: Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the woman's own health is seriously endangered by the pregnancy?
 1. Yes
 2. No
2. **abrape**: Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if she becomes pregnant as a result of rape?
 1. Yes
 2. No
3. **abpoor**: Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the family has a very low income and cannot afford any more children?
 1. Yes
 2. No
4. **abdefect**: Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance

of serious defect in the baby?

1. Yes
2. No

Guns and Crime (2 items)

5. **gunlaw:** Would you favor or oppose a law which would require a person to obtain a police permit before he or she could buy a gun?

1. Favor
2. Oppose

6. **cappun:** Do you favor or oppose the death penalty for persons convicted of murder?

1. Favor
2. Oppose

Sexual Morality (4 items)

7. **homosex:** What is your opinion about sexual relations between two adults of the same sex—do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?

1. Always wrong
2. Almost always wrong
3. Wrong only sometimes
4. Not wrong at all

8. **premarsx:** If a man and woman have sexual relations before marriage, do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?

1. Always wrong
2. Almost always wrong
3. Wrong only sometimes
4. Not wrong at all

9. **xmarsex:** What is your opinion about a married person having sexual relations with someone other than the marriage partner—is it always wrong, almost always wrong, wrong only sometimes, or not wrong at all?

1. Always wrong
2. Almost always wrong
3. Wrong only sometimes
4. Not wrong at all

10. **pornlaw:** Which of these statements comes closest to your feelings about pornography laws?

1. There should be laws against the distribution of pornography whatever the age
2. There should be laws against the distribution of pornography to persons under 18
3. There should be no laws forbidding the distribution of pornography

Drugs (1 item)

11. **grass:** Do you think the use of marijuana should be made legal or not?

1. Should be legal
2. Should not be legal

Gender Roles (3 items)

12. **fechld**: To what extent do you agree or disagree that: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree

13. **fefam**: Please read the following statement and indicate whether you strongly agree, agree, disagree, or strongly disagree: It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family.

1. Strongly agree
2. Agree
3. Disagree
4. Strongly disagree

14. **fepresch**: Please read the following statement and indicate whether you strongly agree, agree, disagree, or strongly disagree: A preschool child is likely to suffer if his or her mother works.

1. Strongly agree
2. Agree
3. Disagree
4. Strongly disagree

Economic Policy (2 items)

15. **affrmact**: Some people say that because of past discrimination, Black people should be given preference in hiring and promotion. Others say that such preference in hiring and promotion of Black people is wrong because it discriminates against whites. What about your opinion—are you for or against preferential hiring and promotion of Black people?

1. Strongly favor
2. Not strongly favor
3. Not strongly oppose
4. Strongly oppose

16. **eqwlth**: Some people think that the government in Washington ought to reduce the income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing this income difference between the rich and the poor. What score between 1 and 7 comes closest to the way you feel?

1. The government should reduce differences
- 2–6. (Intermediate positions)
7. The government should not concern itself with reducing income differences

Immigration (3 items)

17. **immjobs**: There are different opinions about immigrants from other countries living in America. How much do you agree or disagree with the following statement: Immigrants take jobs away from people who were born in America.

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly

18. **immcrime:** There are different opinions about immigrants from other countries living in America. How much do you agree or disagree with the following statement: Immigrants increase crime rates in America.

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly

19. **immameco:** There are different opinions about immigrants from other countries living in America. How much do you agree or disagree with the following statement: Immigrants are generally good for America's economy.

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly

Free Speech (6 items)

20. **spkath:** There are always some people whose ideas are considered bad or dangerous by other people. For instance, somebody who is against all churches and religion. If such a person wanted to make a speech in your community against churches and religion, should he be allowed to speak, or not?

1. Yes, allowed to speak
2. Not allowed

21. **colath:** There are always some people whose ideas are considered bad or dangerous by other people. For instance, somebody who is against all churches and religion. Should such a person be allowed to teach in a college or university, or not?

1. Yes, allowed to teach
2. Not allowed

22. **spkrac:** Consider a person who believes that Black people are genetically inferior. If such a person wanted to make a speech in your community claiming that Black people are inferior, should he be allowed to speak, or not?

1. Yes, allowed to speak
2. Not allowed

23. **colrac:** Consider a person who believes that Black people are genetically inferior. Should such a person be allowed to teach in a college or university, or not?

1. Yes, allowed to teach

2. Not allowed

24. **spkmslm**: Consider a Muslim clergyman who preaches hatred of the United States. If such a person wanted to make a speech in your community preaching hatred of the United States, should he be allowed to speak, or not?

1. Yes, allowed

2. Not allowed

25. **colmslm**: Consider a Muslim clergyman who preaches hatred of the United States. Should such a person be allowed to teach in a college or university, or not?

1. Yes, allowed to teach

2. Not allowed

Religion (2 items)

26. **god**: Tell me which statement comes closest to expressing what you believe about God.

1. I don't believe in God

2. I don't know whether there is a God and I don't believe there is any way to find out

3. I don't believe in a personal God, but I do believe in a Higher Power of some kind

4. I find myself believing in God some of the time, but not at others

5. While I have doubts, I feel that I do believe in God

6. I know God really exists and I have no doubts about it

27. **bible:** Which of these statements comes closest to describing your feelings about the Bible?

1. The Bible is the actual word of God and is to be taken literally, word for word
2. The Bible is the inspired word of God but not everything in it should be taken literally, word for word
3. The Bible is an ancient book of fables, legends, history, and moral precepts recorded by men

Confidence in Institutions (3 items)

28. **confed:** As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Executive branch of the federal government.

1. A great deal
2. Only some
3. Hardly any

29. **conpress:** As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: The Press.

1. A great deal
2. Only some
3. Hardly any

30. **consci**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: The Scientific Community.
1. A great deal
 2. Only some
 3. Hardly any

Part 2: Non-Culture-War Items (22 questions)

Public Spending Priorities (8 items)

31. **nataid**: We are faced with many problems in this country, none of which can be solved easily or inexpensively. Are we spending too much, too little, or about the right amount on: Foreign aid?
1. Too little
 2. About right
 3. Too much
32. **natspac**: Are we spending too much, too little, or about the right amount on: The space exploration program?
1. Too little
 2. About right
 3. Too much
33. **natsci**: Are we spending too much, too little, or about the right amount on: Supporting scientific research?
1. Too little
 2. About right

3. Too much

34. **natenvir**: Are we spending too much, too little, or about the right amount on: Improving and protecting the environment?

1. Too little

2. About right

3. Too much

35. **nateduc**: Are we spending too much, too little, or about the right amount on: Improving the nation's education system?

1. Too little

2. About right

3. Too much

36. **natheal**: Are we spending too much, too little, or about the right amount on: Improving and protecting the nation's health?

1. Too little

2. About right

3. Too much

37. **natroad**: Are we spending too much, too little, or about the right amount on: Highways and bridges?

1. Too little

2. About right

3. Too much

38. **nat soc**: Are we spending too much, too little, or about the right amount on: Social Security?

1. Too little
2. About right
3. Too much

Institutional Confidence (6 items)

39. **coneduc**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Education.

1. A great deal
2. Only some
3. Hardly any

40. **conbus**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Major companies.

1. A great deal
2. Only some
3. Hardly any

41. **confinan**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Banks and financial institutions.

1. A great deal
2. Only some
3. Hardly any

42. **conlegis**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Congress.

1. A great deal
2. Only some
3. Hardly any

43. **conarmy**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Military.

1. A great deal
2. Only some
3. Hardly any

44. **conmedic**: As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them: Medicine.

1. A great deal
2. Only some
3. Hardly any

Social Trust (3 items)

45. **trust**: Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

1. Can trust
2. Can't be too careful

3. Depends

46. **fair:** Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

1. Would take advantage of you

2. Would try to be fair

3. Depends

47. **helpful:** Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

1. Try to be helpful

2. Looking out for themselves

3. Depends

Economic Outlook and Class (3 items)

48. **satfin:** We are interested in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all?

1. Pretty well satisfied

2. More or less satisfied

3. Not satisfied at all

49. **finalter:** During the last few years, has your financial situation been getting better, worse, or has it stayed the same?

1. Getting better

2. Stayed the same

3. Getting worse

50. **getahead**: Some people say that people get ahead by their own hard work; others say that lucky breaks or help from other people are more important. Which do you think is most important?

1. Hard work most important

2. Hard work, luck equally important

3. Luck most important

Work Centrality (2 items)

51. **richwork**: If you were to get enough money to live as comfortably as you would like for the rest of your life, would you continue to work or would you stop working?

1. Continue to work

2. Stop working

52. **satjob**: On the whole, how satisfied are you with the work you do—would you say you are very satisfied, moderately satisfied, a little dissatisfied, or very dissatisfied?

1. Very satisfied

2. Moderately satisfied

3. A little dissatisfied

4. Very dissatisfied

Table 1: Large language models included in the analysis.

Provider	Model	OpenRouter Identifier
AI21	Jamba Large 1.7	ai21/jamba-large-1.7
Allen AI	OLMo 3 7B Instruct	allenai/olmo-3-7b-instruct
Anthropic	Claude 3.7 Sonnet	anthropic/claude-3.7-sonnet
Anthropic	Claude Sonnet 4.5	anthropic/claude-sonnet-4.5
Cohere	Command R+ (08-2024)	cohere/command-r-plus-08-2024
DeepCogito	Cogito v2 Preview (Llama 405B)	deepcogito/cogito-v2-preview-llama-405b
DeepSeek	DeepSeek Chat v3 (0324)	deepseek/deepseek-chat-v3-0324
DeepSeek	DeepSeek Chat v3.1	deepseek/deepseek-chat-v3.1
DeepSeek	DeepSeek v3.2 (20251201)	deepseek/deepseek-v3.2-20251201
Google	Gemini 2.5 Flash	google/gemini-2.5-flash
Google	Gemini 2.5 Flash Lite	google/gemini-2.5-flash-lite
Google	Gemma 3 12B IT	google/gemma-3-12b-it
Meta	Llama 3.1 8B Instruct	meta-llama/llama-3.1-8b-instruct
Meta	Llama 3.3 70B Instruct	meta-llama/llama-3.3-70b-instruct
Meta	Llama 4 Maverick	meta-llama/llama-4-maverick
Mistral AI	Mistral Large (2512)	mistralai/mistral-large-2512
Mistral AI	Mistral Medium 3.1	mistralai/mistral-medium-3.1
Mistral AI	Mistral Nemo	mistralai/mistral-nemo
Mistral AI	Mistral Small 3.2 24B Instruct	mistralai/mistral-small-3.2-24b-instruct
Moonshot AI	Kimi K2	moonshotai/kimi-k2
OpenAI	GPT-4o Mini	openai/gpt-4o-mini
OpenAI	GPT-5	openai/gpt-5
OpenAI	GPT-5 Mini	openai/gpt-5-mini
OpenAI	GPT-OSS 120B	openai/gpt-oss-120b
Qwen	Qwen 2.5 72B Instruct	qwen/qwen-2.5-72b-instruct
Qwen	Qwen3 235B A22B (2507)	qwen/qwen3-235b-a22b-2507
xAI	Grok 4 Fast	x-ai/grok-4-fast
Zhipu AI	GLM 4.6	z-ai/glm-4.6

I Prompt format

It is now {year}. You are answering survey questions as the following person, who is living in the United States:

{persona}

Question: {question}

Options:
{options}

Respond with ONLY the number of your answer (e.g., "1" or "2"). Do not explain your reasoning.

Figure 30: Prompt template for LLM attitude elicitation. Prompts were sent as a single user message with no system message, temperature set to 1.0, and maximum output length of 50 tokens. {persona} is a natural language biographical description constructed from the respondent's GSS demographic record (see Figure 31). {options} are numbered response categories formatted as 1. Option text, one per line.

It is now 2024. You are answering survey questions as the following person, who is living in the United States:

I am 42 years old. I am female. I am White. I am married. I completed 4 years of college (bachelor's degree). My family income is \$75,000 to \$89,999. I am a not very strong Democrat. I am slightly liberal. My religion is Protestant. I attend religious services about once a month.

Question: Would you favor or oppose a law which would require a person to obtain a police permit before he or she could buy a gun?

Options:
1. Favor
2. Oppose

Respond with ONLY the number of your answer (e.g., "1" or "2"). Do not explain your reasoning.

Figure 31: Example of a fully instantiated prompt.