Appendix: "What's a Parent to Do? Measuring Cultural Logics of Parenting with Computational Text Analysis"

Replication files

The code for the analysis in this paper and in this appendix are available at:
https://github.com/ophastings/replications/tree/main/2024-parenting-logics-text-analysis-SSR

More detail on the open-ended questions

The data and survey materials (including the full questionnaire) are publicly available at https://osf.io/bbehq/. The specific question which gathered the open-ended question is as follows:

> We're interested in how people think about what it means to be a good parent. We'll describe 2 parenting situations on the pages that follow. Each situation involves a parent with a child between the ages of 8 and 10.
>
> For the first situation, we'll ask how you would advise the parent to respond. For the second situation, we'll describe how a parent behaved, and will ask what you think of that behavior and why. Please read each description carefully and be as thorough as possible in your responses. Your answers are extremely important to us.
>
> [The vignette was presented here, with a randomly assigned parent name and child gender]
>
> How would you advise [parent's name] to respond in this situation?

Less than 3% of respondents left Question 1 blank. The average number of words and interquartile range is shown in the table below by scenario.

Scenario	Mean words	Median Words	IQ Range (25th-75th percentile)
Bored at School	20.2	15	9-26
Bedtime Rule	25.3	23	13-38
Take Your Medicine	22.5	19	11-33

Those with at least a BA gave slightly longer responses (mean = 28.2 words), compared to those with Some College (25.5 words) or a High School Degree or less (21.0 words). There was less variation in response length by race/ethnicity, ranging from 22.3 words (Hispanic) to 26.4 (Whites). Females gave slighlty longer answer than males (mean of 27.3 vs 22.7 words, respectively). All responses were included in the topic models, although dropping all responses with 3 or fewer words had almost no impact on the results.

Selecting the number of topics (k)

Semantic coherence and exclusivity are two common metrics to assess the ideal number of topics emerging from topic models. Semantic coherence scores a single topic by measuring the degree of semantic similarity between high-scoring words in the topic (Mimno et al. 2011). This measure is maximized when the most probable words in a given topic frequently co-occur together, and it's a metric that correlates well with human judgement of topic quality.[1] The exclusivity measure is calculated for a topic within a model by summing up the score of a frequency/exclusivity metric for the most recurring words within the topic itself. For each parenting situation, we created BTM models with different values of k and plotted their mean coherence and exclusivity values in Figure A1.

Scholars caution that while these quantitative metrics can offer some guidance about what values of k are on the "coherence-exclusivity frontier" (models in the upper-right quadrant are better), they do not necessarily provide a recommendation as to the single best model. It is important to include human evaluation when choosing topics, as ultimately topics need to be useful, and the topics most interpretable to humans are not necessarily those that perform best using quantitative metrics for model selection (Wallach et al. 2009; Roberts et al. 2014; Grimmer et al. 2022). Moreover, in our context there may not even be a "best" k for any particular situation, nor is it necessarily the case that each scenario should have the same number of topics. Models with a few topics might better capture generalized parenting logics, while models with many topics may capture specific nuances within topics.

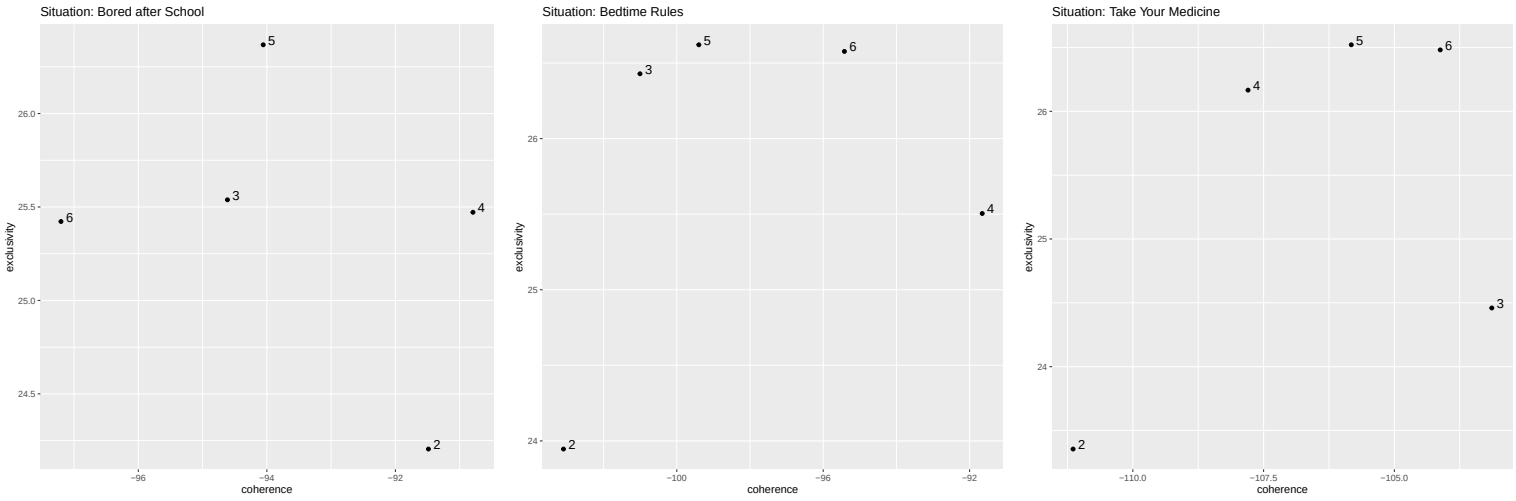

Figure A1: Comparing Exclusivity and Coherence for Different Numbers of Topics (k)

Figure A1 suggests that the two-topic model performs very poorly in terms of both semantic coherence and exclusivity, aligning with our human-based assessment that the two-topic model does not well capture heterogeneity in the data. We found that we needed at least four topics to reach the "coherence-exclusivity frontier," and adding more topics only marginally improved the fit of the models (including beyond the k = 6 that we show in the plots). We also found that when using much larger values for k, the topics were less interpretable. Theoretically, this analysis suggests that our data provide evidence of a set of parenting logics that do not necessarily align with the dichotomous categorization of parenting styles discussed in the aforementioned research. Another approach might be to generate a much larger number of topics. Some of these topics might capture different nuances of similar parenting logics and could be combined, while others might capture aspects that are less related to parenting logics. This approach would seem more promising in a scenario with a much larger number of responses, where each topic could then be represented by more responses.

Fractional Multinomial Logistic (FML) models

FML regression models are an extension of the fractional logit model (Papke and Wooldridge 1996). Like the multinomial logistic regression model (mlogit), FML estimates the associations with a single dependent variable consisting of multiple categories. However, mlogit assumes that the categories are mutually exclusive, meaning an observation can belong to one category (=1) but not to the other categories (=0). FML fits by quasi maximum likelihood a fractional multinomial logit model. Each variable ranges between 0 and 1 and all variables must, for each observation, add up to 1. For example, they may be proportions (and most commonly they are). Fractional models have been widely used in social science research, including to study the determinants of Brexit votes in the UK (Alaimo and Solivetti 2019), the relationship between local violence within a neighborhood and time spent at home in the US (Browning et al. 2021), public good provision in India (Roy Chowdhury 2021), elderly care provision in China (Hu and Chen 2019), and the division of labor within households in Austria (Spitzer and Hammer 2016)

Table A1 presents the full FML models underlying Figure 3 in the paper. In these models, the baseline topic is the first topic, i.e., the most prevalent (BS-1, BR-1, and TM-1, respectively). Coefficients and standard errors refer to the association between that specific variable and a specific topic relative to the association between that same variable and the first topic. The remaining tables show models excluding race/ethnicity, education, and income.

Table A1: Fractional Multinomial Logit Estimates, by Parenting Situation

	Fractional Multinomial Logit (Ref.: Topic 1)								
	Parenting Situation 1			**Parenting Situation 2**			**Parenting Situation 3**		
	BS-2	BS-3	BS-4	BR-2	BR-3	BR-4	TM-2	TM-3	TM-4
Income quartile (Ref.: Quartile 1)									
Quartile 2	0.430*	0.121	-0.072	0.377	-0.109	0.066	0.323	0.334	0.646*
	(0.221)	(0.332)	(0.323)	(0.289)	(0.287)	(0.317)	(0.220)	(0.286)	(0.338)
Quartile 3	0.216	0.302	-0.332	0.170	0.028	0.072	0.488*	0.198	0.403
	(0.256)	(0.381)	(0.367)	(0.264)	(0.297)	(0.310)	(0.280)	(0.355)	(0.378)
Quartile 4	0.214	-0.288	-0.173	-0.132	-0.157	-0.296	0.421	0.315	0.233
	(0.256)	(0.404)	(0.385)	(0.331)	(0.369)	(0.396)	(0.280)	(0.331)	(0.369)
Education (Ref.: High school or less)									
Some college	0.015	0.652*	0.492*	0.051	0.109	0.667**	0.142	0.294	0.160
	(0.196)	(0.348)	(0.285)	(0.244)	(0.255)	(0.298)	(0.212)	(0.281)	(0.316)
Bachelor's degree or higher	-0.030	0.056	0.237	-0.134	0.230	0.597*	-0.048	0.535	0.049
	(0.215)	(0.393)	(0.282)	(0.239)	(0.268)	(0.326)	(0.250)	(0.332)	(0.317)
Race/Ethnicity (Ref.: White, Non-Hispanic)									
Black, Non-Hispanic	-0.418	-0.338	0.242	-0.573	-0.618*	-0.464	-0.239	-0.812***	-0.236
	(0.263)	(0.418)	(0.427)	(0.363)	(0.335)	(0.388)	(0.261)	(0.246)	(0.369)
Other, Non-Hispanic	-0.458	-0.059	-0.332	-0.447	0.215	0.090	0.236	0.238	0.401
	(0.319)	(0.522)	(0.396)	(0.278)	(0.395)	(0.398)	(0.409)	(0.536)	(0.582)
Hispanic	-0.594**	-0.727**	0.051	0.214	-0.060	0.267	0.102	-0.128	0.011
	(0.240)	(0.330)	(0.326)	(0.236)	(0.256)	(0.302)	(0.217)	(0.302)	(0.399)
Female	0.319**	-0.456*	-0.508**	-0.021	0.162	0.160	0.224	0.194	-0.290
	(0.159)	(0.270)	(0.256)	(0.174)	(0.184)	(0.220)	(0.178)	(0.217)	(0.260)
Age	-0.002	-0.014	0.019	0.030**	0.030**	0.036***	0.010	-0.020*	0.009
	(0.009)	(0.013)	(0.013)	(0.013)	(0.012)	(0.014)	(0.009)	(0.011)	(0.011)
Parent: Mother	-0.224	-0.024	-0.111	-0.235	-0.132	-0.200	0.006	0.264	-0.287
	(0.154)	(0.234)	(0.238)	(0.170)	(0.175)	(0.202)	(0.166)	(0.206)	(0.252)
Child: Son	0.271*	0.364	0.042	0.102	0.123	0.193	-0.098	0.047	0.048
	(0.151)	(0.247)	(0.235)	(0.171)	(0.176)	(0.206)	(0.167)	(0.206)	(0.247)
Constant	-0.418	-1.272**	-2.048***	-1.049*	-1.571***	-2.600***	-1.328***	-1.076*	-2.143***
	(0.442)	(0.640)	(0.615)	(0.545)	(0.472)	(0.617)	(0.498)	(0.590)	(0.534)
Observations		578			551			535	

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table A2: Fractional Multinomial Logit Estimates, by Parenting Situation, Excluding Race/Ethnicity

	Fractional Multinomial Logit (Ref: Topic 1)								
	Parenting Situation 1			**Parenting Situation 2**			**Parenting Situation 3**		
	BS-2	BS-3	BS-4	BR-2	BR-3	BR-4	TM-2	TM-3	TM-4
Income quartile (Ref.: Quartile 1)									
Quartile 2	0.571***	0.239	-0.102	0.408	-0.039	0.094	0.338	0.426	0.668**
	(0.218)	(0.336)	(0.348)	(0.305)	(0.299)	(0.322)	(0.222)	(0.290)	(0.319)
Quartile 3	0.320	0.375	-0.367	0.236	0.149	0.127	0.518*	0.316	0.444
	(0.251)	(0.373)	(0.385)	(0.258)	(0.276)	(0.294)	(0.279)	(0.363)	(0.367)
Quartile 4	0.347	-0.195	-0.212	-0.092	-0.047	-0.256	0.445	0.464	0.270
	(0.260)	(0.404)	(0.400)	(0.327)	(0.362)	(0.396)	(0.279)	(0.350)	(0.370)
Education (Ref.: High school or less)									
Some college	0.001	0.677**	0.461	-0.072	0.057	0.574*	0.071	0.170	0.118
	(0.199)	(0.335)	(0.284)	(0.244)	(0.254)	(0.297)	(0.200)	(0.262)	(0.309)
Bachelor's degree or higher	0.057	0.226	0.178	-0.216	0.252	0.536*	-0.067	0.529*	0.073
	(0.213)	(0.354)	(0.264)	(0.250)	(0.272)	(0.318)	(0.241)	(0.315)	(0.338)
Female	0.327**	-0.439	-0.497*	-0.072	0.111	0.113	0.235	0.199	-0.286
	(0.159)	(0.276)	(0.257)	(0.177)	(0.186)	(0.218)	(0.180)	(0.220)	(0.254)
Age	0.001	-0.011	0.018	0.028**	0.028**	0.035**	0.008	-0.023**	0.008
	(0.009)	(0.013)	(0.013)	(0.014)	(0.012)	(0.014)	(0.009)	(0.011)	(0.011)
Parent: Mother	-0.214	-0.034	-0.098	-0.239	-0.128	-0.195	0.009	0.266	-0.286
	(0.153)	(0.238)	(0.239)	(0.173)	(0.180)	(0.206)	(0.167)	(0.208)	(0.251)
Child: Son	0.280*	0.352	0.056	0.124	0.119	0.217	-0.103	0.030	0.038
	(0.152)	(0.249)	(0.238)	(0.176)	(0.180)	(0.209)	(0.167)	(0.206)	(0.250)
Constant	-0.882**	-1.708***	-1.957***	-0.991*	-1.629***	-2.513***	-1.266***	-1.118**	-2.119***
	(0.412)	(0.605)	(0.633)	(0.572)	(0.490)	(0.612)	(0.478)	(0.557)	(0.528)
Observations	578	578	578	551	551	551	535	535	535

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table A3: Fractional Multinomial Logit Estimates, by Parenting Situation, Excluding Income

	Fractional Multinomial Logit (Ref.: Topic 1)								
	Parenting Situation 1			**Parenting Situation 2**			**Parenting Situation 3**		
	BS-2	BS-3	BS-4	BR-2	BR-3	BR-4	TM-2	TM-3	TM-4
Education (Ref.: High school or less)									
Some college	0.052	0.706**	0.429	0.064	0.094	0.657**	0.196	0.333	0.229
	(0.189)	(0.330)	(0.312)	(0.231)	(0.254)	(0.290)	(0.212)	(0.276)	(0.309)
Bachelor's degree or higher	-0.015	0.011	0.137	-0.162	0.205	0.541*	0.059	0.593**	0.044
	(0.196)	(0.363)	(0.296)	(0.213)	(0.245)	(0.301)	(0.227)	(0.290)	(0.300)
Race/Ethnicity (Ref.: White, Non-Hispanic)									
Black, Non-Hispanic	-0.532**	-0.389	0.294	-0.609*	-0.611*	-0.467	-0.335	-0.883***	-0.302
	(0.256)	(0.415)	(0.453)	(0.332)	(0.312)	(0.366)	(0.259)	(0.254)	(0.374)
Other, Non-Hispanic	-0.475	-0.082	-0.319	-0.467*	0.215	0.081	0.231	0.224	0.384
	(0.310)	(0.530)	(0.391)	(0.274)	(0.402)	(0.399)	(0.413)	(0.538)	(0.604)
Hispanic	-0.659***	-0.742**	0.042	0.192	-0.056	0.269	0.042	-0.180	-0.020
	(0.241)	(0.328)	(0.333)	(0.236)	(0.244)	(0.303)	(0.224)	(0.310)	(0.396)
Female	0.312*	-0.448*	-0.487*	-0.054	0.163	0.146	0.195	0.165	-0.349
	(0.160)	(0.264)	(0.249)	(0.175)	(0.188)	(0.223)	(0.177)	(0.213)	(0.271)
Age	-0.002	-0.013	0.017	0.028**	0.029***	0.034***	0.012	-0.018*	0.008
	(0.009)	(0.012)	(0.013)	(0.011)	(0.011)	(0.013)	(0.009)	(0.011)	(0.010)
Parent: Mother	-0.225	-0.009	-0.117	-0.230	-0.128	-0.190	0.005	0.263	-0.270
	(0.154)	(0.234)	(0.241)	(0.169)	(0.178)	(0.203)	(0.168)	(0.204)	(0.251)
Child: Son	0.264*	0.317	0.063	0.091	0.130	0.194	-0.062	0.069	0.059
	(0.153)	(0.250)	(0.238)	(0.169)	(0.176)	(0.207)	(0.168)	(0.208)	(0.250)
Constant	-0.184	-1.199**	-2.087***	-0.807	-1.583***	-2.481***	-1.160**	-0.900	-1.706***
	(0.441)	(0.580)	(0.594)	(0.520)	(0.475)	(0.624)	(0.463)	(0.565)	(0.576)
Observations	578	578	578	551	551	551	535	535	535

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table A4: Fractional Multinomial Logit Estimates, by Parenting Situation, Excluding Education

	Fractional Multinomial Logit (Ref.: Topic 1)								
	Parenting Situation 1			Parenting Situation 2			Parenting Situation 3		
	BS-2	BS-3	BS-4	BR-2	BR-3	BR-4	TM-2	TM-3	TM-4
Income quartile (Ref.: Quartile 1)									
Quartile 2	0.431**	0.188	0.005	0.362	-0.059	0.207	0.336	0.391	0.667**
	(0.217)	(0.327)	(0.330)	(0.274)	(0.285)	(0.321)	(0.222)	(0.288)	(0.326)
Quartile 3	0.207	0.378	-0.197	0.137	0.108	0.273	0.480*	0.334	0.420
	(0.235)	(0.349)	(0.392)	(0.239)	(0.278)	(0.299)	(0.274)	(0.329)	(0.378)
Quartile 4	0.197	-0.335	-0.070	-0.198	-0.035	-0.036	0.391	0.519*	0.245
	(0.239)	(0.378)	(0.392)	(0.300)	(0.333)	(0.365)	(0.256)	(0.289)	(0.338)
Race/Ethnicity (Ref.: White, Non-Hispanic)									
Black, Non-Hispanic	-0.420	-0.356	0.234	-0.544	-0.629*	-0.390	-0.190	-0.749***	-0.184
	(0.261)	(0.420)	(0.428)	(0.356)	(0.342)	(0.398)	(0.260)	(0.249)	(0.348)
Other, Non-Hispanic	-0.463	-0.023	-0.276	-0.454	0.233	0.167	0.223	0.367	0.412
	(0.319)	(0.471)	(0.384)	(0.277)	(0.379)	(0.401)	(0.412)	(0.529)	(0.593)
Hispanic	-0.586**	-0.670**	0.041	0.241	-0.110	0.146	0.093	-0.226	-0.009
	(0.232)	(0.324)	(0.332)	(0.234)	(0.252)	(0.286)	(0.209)	(0.289)	(0.403)
Female	0.320**	-0.361	-0.435*	-0.023	0.173	0.198	0.231	0.213	-0.280
	(0.161)	(0.258)	(0.254)	(0.179)	(0.186)	(0.221)	(0.180)	(0.222)	(0.260)
Age	-0.002	-0.012	0.020	0.030**	0.029**	0.035**	0.009	-0.018	0.009
	(0.009)	(0.013)	(0.013)	(0.013)	(0.012)	(0.014)	(0.009)	(0.011)	(0.011)
Parent: Mother	-0.226	-0.027	-0.103	-0.243	-0.133	-0.214	0.004	0.276	-0.287
	(0.154)	(0.239)	(0.236)	(0.170)	(0.175)	(0.203)	(0.167)	(0.206)	(0.253)
Child: Son	0.270*	0.352	0.032	0.097	0.125	0.196	-0.092	0.037	0.050
	(0.151)	(0.249)	(0.237)	(0.171)	(0.176)	(0.207)	(0.167)	(0.208)	(0.248)
Constant	-0.420	-1.197*	-1.994***	-1.067**	-1.487***	-2.246***	-1.289***	-0.958*	-2.094***
	(0.441)	(0.642)	(0.623)	(0.521)	(0.441)	(0.607)	(0.484)	(0.570)	(0.523)
Observations	578	578	578	551	551	551	535	535	535

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Other Supplemental Models

We also examined models where we pooled the topics across vignettes by creating variables for "any pedagogic", "any pragmatic", "any negotiated," ' and "any assertive" based on our topic labels. Each variable is binary, so we ran logistic regression models predicting the probability for each variable. These results are shown below in Table A5. Despite the increased sample size in each model (N = 1,664), none of the coefficients for race/ethnicity, education, or income are significant at the p <.05 level. We further show in Figure A2 the predicted probability of each of these outcomes across sociodemographic categories. As with the results in the paper, these show relatively flat lines, suggesting there are not huge average differences between groups, and that there is substantial variation within them. This corroborates our main findings.

Table A5: Logistic Regression Coefficients Predicting "any pedagogic", "any pragmatic", "any negotiated," ' and "any assertive" based on topic labels

	(1) pedagogic	(2) pragmatic	(3) negotiated	(4) assertive
Income (ref = Quartile 1)				
Income quartile 2	-0.32+	0.25	0.0030	-0.0055
	(0.18)	(0.18)	(0.23)	(0.19)
Income quartile 3	-0.18	0.29	-0.28	-0.20
	(0.20)	(0.20)	(0.24)	(0.22)
Income quartile 4	-0.31	0.25	-0.028	0.077
	(0.24)	(0.24)	(0.28)	(0.25)
Education (ref = High school or less)				
Some college	0.085	0.011	0.19	-0.073
	(0.16)	(0.16)	(0.21)	(0.18)
BA+	0.087	-0.16	0.37+	0.25
	(0.18)	(0.18)	(0.22)	(0.19)
Race/ethnicity (ref = White)				
Black	0.17	-0.11	-0.22	0.061
	(0.20)	(0.21)	(0.25)	(0.22)
Other	-0.20	0.45+	0.051	0.29
	(0.26)	(0.25)	(0.35)	(0.26)
Hispanic	0.12	-0.022	-0.085	0.20
	(0.18)	(0.19)	(0.23)	(0.20)
Female	-0.020	0.12	-0.22	-0.21
	(0.13)	(0.13)	(0.16)	(0.14)
Age	0.016*	-0.0039	0.0044	-0.0026
	(0.0070)	(0.0072)	(0.0078)	(0.0077)
Parent is a mother	-0.017	-0.21+	-0.051	-0.091
	(0.12)	(0.12)	(0.15)	(0.13)
Child is a son	-0.036	0.076	-0.074	-0.024
	(0.12)	(0.12)	(0.15)	(0.13)
Constant	-0.71*	-0.24	-1.57***	-0.55
	(0.36)	(0.36)	(0.40)	(0.39)
Observations	1664	1664	1664	1664

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

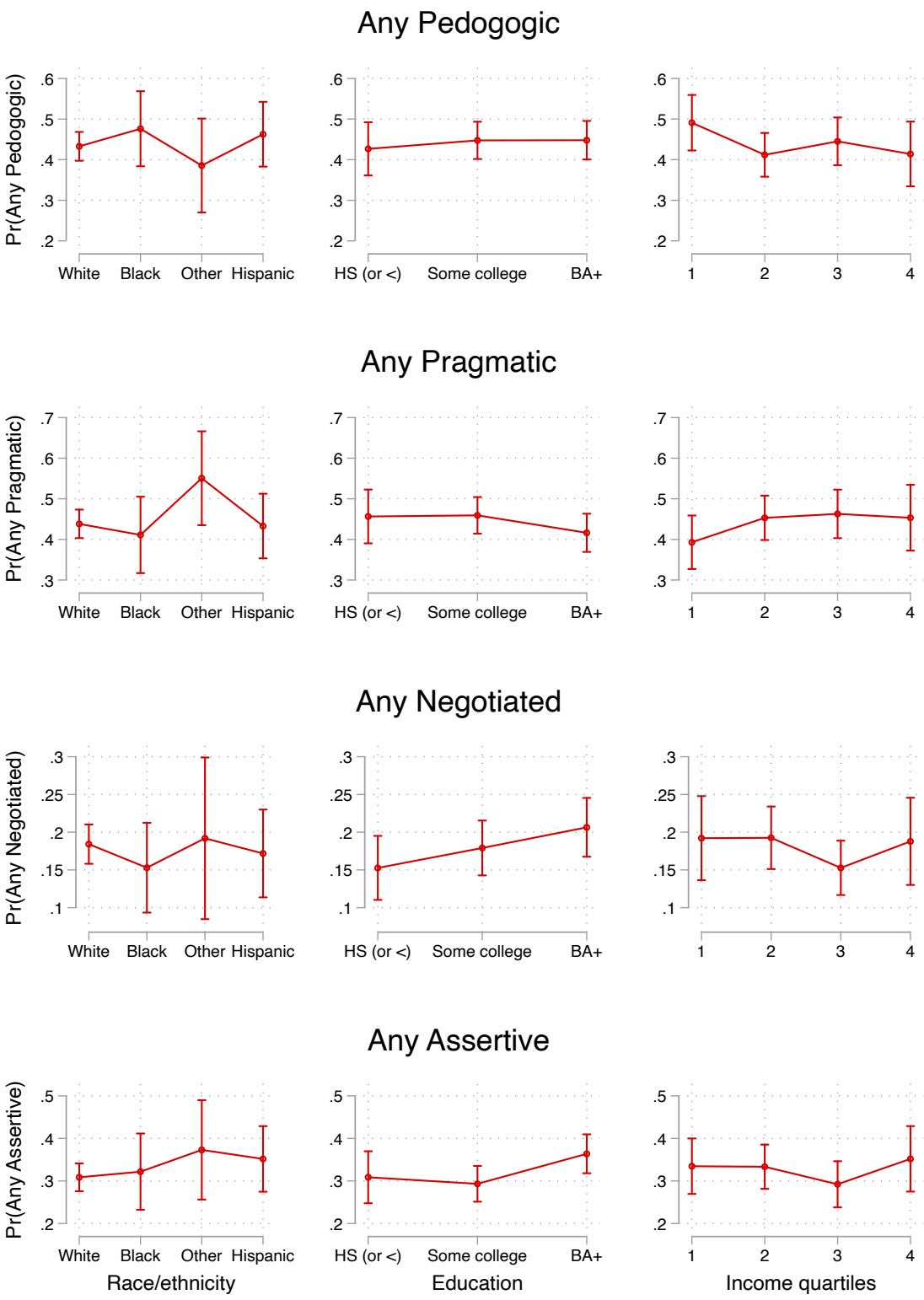

Figure A2: Predicted Probabilities based on Table A5

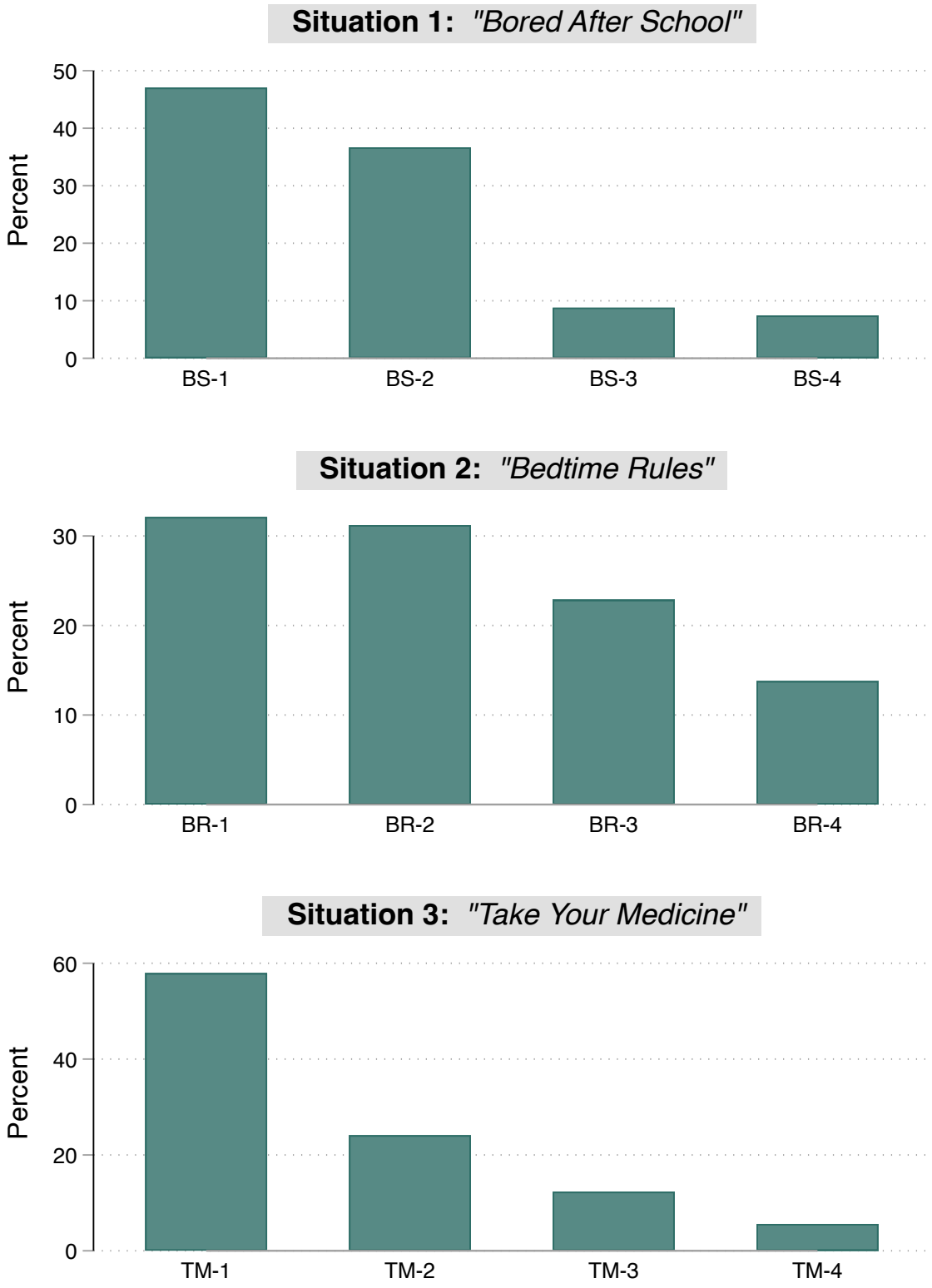

Figure A3: Distribution of Topics in Each Parenting Situation: "Bored After School" (top panel), "Bedtime Rules" (middle panel), and "Take Your Medicine" (bottom panel)

Appendix References

Alaimo, Leonardo S. and Luigi M. Solivetti. 2019. "Territorial Determinants of the Brexit Vote." *Social Indicators Research* 144(2):647–667.

Browning, Christopher R., Catherine A. Calder, Bethany Boettner, Jake Tarrence, Kori Khan, Brian Soller, and Jodi L. Ford. 2021. "Neighborhoods, Activity Spaces, and the Span of Adolescent Exposures." *American sociological review* 86(2):201–233.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Hu, Anning and Feinian Chen. 2019. "Which Child Is Parents' Preferred Caregiver/Listener in China?" *Research on Aging* 41(4):390–414.

Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 262–272.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401 (k) Plan Participation Rates." *Journal of applied econometrics* 11(6):619–632.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Roy Chowdhury, Joyita. 2021. "Elite Status, Market Linkages, and Contributions to Collective Goods: Evidence from a Survey and Public Goods Experiments." *The Journal of Development Studies* pp. 1–17.

Spitzer, Sonja and Bernhard Hammer. 2016. "The Division of Labour Within Households: Fractional Logit Estimates Based on the Austrian Time Use Survey." *AGENTA Working Paper, No 4/2016* .

Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. "Evaluation Methods for Topic Models." In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1105–1112.