Online Appendix to
Ideology and compliance with health guidelines
during the COVID-19 pandemic: A comparative
perspective

Contents

A. Survey fieldwork 1

B. Non-compliance list experiment 2
   B.1. Social distancing policies 2
   B.2. Relationship between non-compliance and social ties 2
   B.3. Exploring experimental design assumptions 3
   B.4. Sample characteristics at baseline for treatment and control units 5

C. Latent variable model behavioral health changes 7
   C.1. Random coefficient hierarchical factor model 9
   C.2. Estimates and comparison to list experiment 10

D. Nonparametric estimates details 14
   D.1. Model and estimates 14
   D.2. Accounting for government support 18
   D.3. Accounting for trust in experts 19
   D.4. Model comparisons 21
   D.5. Comparisons to pooled linear fixed effects models 21
A. Survey fieldwork

Background: All countries included in the analysis had numerous confirmed COVID-19 cases and all had reported COVID-19 related deaths at the time of the survey (see Table A.1). Facing the same pandemic, governments had put in place new health guidelines that emphasized the importance of social distancing to reduce the spread of the virus, alongside other behavioral changes, such as more frequent and thorough hand-washing. Across countries, the general governmental recommendation was not to meet other people and stay home whenever possible. For example, in France all public and private gatherings were banned and in Germany the federal government declared that ‘rule number 1’ was to reduce social contact to a minimum. The US president declared a national emergency on March 13, 2020, and in most US states stay-at-home-orders were in place during the time of the survey (with more than 90% of the population being confined or partially confined according to our data). The exception is Sweden. While Swedish public health authorities also emphasized that everyone has a personal responsibility to prevent transmission and discouraged large events, they did not generally recommend social distancing except for older people.

The surveys were in the field between April 15 and April 20 2020 carried out by IPSOS in Austria, France, Germany, Italy, New Zealand, Sweden, and the United Kingdom. In Australia and the United States data collection was conducted by CSA Research. Table A.1 lists fieldwork periods, sample sizes, and the survey completion rate of participating respondents in each country. Note that variation in sample sizes reflects resource constraints unrelated to the analysis as noted in our pre-analysis plan. The last two columns includes two macro variables: The median number of deaths ascribed to Covid-19 in the week prior to the survey as share of the total population and an index of the stringency of the overall governmental response from the Oxford Covid-19 Government Response Tracker.

Sampling was done as part of ongoing online panels using quota sampling. Dropout rates are relatively low. In most countries more than 90% of respondents completed the survey after agreeing to participate. In Australia and the United States, the completion rate is closer to 70%. The resulting samples were weighted by the survey providers to match Census population margins for gender, age, occupation, region, and degree of urbanization (the latter was not used in New Zealand).

---

1The data were collected for the collaborative project “Citizens’ Attitudes Under COVID-19 Pandemic” by the following research team: Sylvain Brouard (Sciences Po, CEVIPOF & LIEPP), Michael Becher (IAST-Université Toulouse Capitole 1), Martial Foucault (Sciences Po-CEVIPOF), Pavlos Vasilopoulos (University of York), Vincenzo Galasso (Bocconi University), Christoph Hönnige (University of Hanover), Eric Kerrouche (Sciences Po-CEVIPOF), Vincent Pons (Harvard Business School), Hanspeter Kriesi (EUI), Richard Nadeau (University of Montreal), Dominique Reynié (Sciences Po-CEVIPOF), and Daniel Stegmueller (Duke University).

2In our list experimental analyses, we excluded 32 cases (0.266%) with excessively large weights (> 5). Results have been replicated with these cases included as well.
B. Non-compliance list experiment

B.1. Social distancing policies

Figure B.1 shows the specific rules on social distancing at the time when the surveys were in the field.

B.2. Relationship between non-compliance and social ties

In this section we explore if country patterns of physically meeting friends and relatives during the pandemic are simply a product of the intensity of existing social ties in a country.

Figure B.1 shows that our list experiment is not contaminated by country-differences in the strength of social or family ties. It plots the share of the population in each country not following social distancing, estimated from the list experiment, against pre-COVID-19 social connections measured by the average time spent socializing with friends and family. Specifically, we use data from the OECD (2020: Figure 11.3) on time (in hours) spent per week interacting with family and friends as a primary activity calculated from Eurostat’s Harmonised European Time Use Surveys (from 2018 or previous years). Figure B.1 illustrates that pre-pandemic patterns of socializing are not strongly related to the share of individuals not following health guidelines during the pandemic. The Spearman rank correlation of socializing and non-compliance with social distancing is 0.02 with a p-value of 0.98.
### Table B.1
Rules on social distancing during fieldwork

| Country   | Rules on social distancing                                                                 |
|-----------|------------------------------------------------------------------------------------------|
| Australia | National restrictions on public gathering and requirement not to leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. The Prime Minister advised against gatherings of more than two people. States states implement specific restrictions (with strict stay-at-home order in most populous state (NSW)) |
| Austria   | Gatherings in public spaces prohibited and requirement not to leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. |
| France    | General ban on gatherings during lockdown and requirement not to leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. |
| Germany   | Ban on both public and private assemblies of more than 2 people from different households by Bund and Länder. Chancellor told people not to visit friends and family |
| Italy     | General ban on gatherings during lockdown and requirement not to leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. |
| New Zealand | Ban on public gatherings and people are instructed to stay at home during alert level 4, except for daily exercise, grocery shopping, and ‘essential’ trips. |
| Sweden    | Ban on public gatherings with more than 50 people but not on meeting friends and family. Personal responsibility for private events. |
| UK        | Ban on public gatherings of more than 2 and visits to friends/family; requirement not to leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. |
| US        | Restrictions on gatherings. Rules vary between states. In 43 states there was some form of stay-at-home-orders active, with requirement no to leave house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips. |

Sources: Oxford Covid-19 Government Response Tracker (Hale et al. 2020) and authors’ corroboration via official and media sources.

### B.3. Exploring experimental design assumptions

The first two columns of Table B.2 shows average item counts in the control group (as well as the coefficient of variation) by country. They indicate that ceiling effects are not a likely concern. In all countries the control group mean item count is below 1.5 with a coefficient of variation around one. However, observing responses close to zero raises the potential issue that a large fraction of respondents choose the rational strategy of replying with ‘0’ simply to ensure that there is no chance that they can be associated with a social norm violation. Column $Y_0$ and $Y_1$ of Table B.2 reports the fraction of respondents reporting having committed none of the acts in the list presented to them for the control and treatment group, respectively. If many respondents indeed follow a rational ‘0’ strategy, we would expect to find that the fraction of ‘0’ responses to be considerably higher in the treated group (who do see the norm violation item) than in to the control group. But, while we do find a sizable share of ‘0’ respondents in the control group, the corresponding share in the treatment group is generally the same or lower. These results suggest that those exposed to the norm violation
treatment are not more likely to shift to a strategy of ‘0’ responses. The exception to this pattern is the United Kingdom, where we find that the fraction of ‘0’ responses among the treated is 6 percentage points higher than among the control group.

Blair and Imai (2012) provide a more sophisticated test of possible design effects in list experiments. A design effect occurs when responses to the control items change due to the presence of the norm violating item. This might be due to respondents evaluating items relative to each other, emotional responses induced by the presence of a sensitive item, or the rational ‘0’ strategy discussed above. The final two columns of Table B.2 shows p values for tests of the null hypothesis of no design effect. The column labelled $p_{BH}$ additionally adjusts $p$ values for multiple country tests using the false-discovery rate controlling procedure of BH. The results clearly do not indicate the presence of design effects in 8 out of 9 experiments: we cannot reject the null hypothesis of no design effect in all countries except the United Kingdom. In the United Kingdom the statistical detection of design effects depends on the decision to adjust for multiple comparisons. Thus, results for the UK should at least be treated with caution. We therefore ensured that excluding the United Kingdom does not affect our

---

3Note that the Blair Imai test already Bonferroni-adjusts $p$ values for multiple testing within countries (Blair and Imai 2012: 64).
Table B.2
Size of control and treatment group, item counts in control group (means and coefficients of variation), proportion of zeros in control and treatment group, and test for design effect.

| Country       | N₀   | N₁   | Y₀ avg. | Y₀ CV | Prop. zeros | Y₀ | Y₁ | Design effect |
|---------------|------|------|---------|-------|-------------|----|----|---------------|
| Australia     | 513  | 494  | 1.395   | 0.823 | 0.23        | 0.22|    | 1.000         |
| Austria       | 498  | 498  | 1.061   | 1.014 | 0.37        | 0.27|    | 0.915         |
| France        | 1014 | 1006 | 0.696   | 1.187 | 0.48        | 0.45|    | 0.984         |
| Germany       | 1000 | 1000 | 1.081   | 1.009 | 0.36        | 0.12|    | 0.098         |
| Italy         | 498  | 499  | 0.753   | 1.318 | 0.53        | 0.54|    | 0.616         |
| New Zealand   | 500  | 498  | 1.131   | 0.681 | 0.16        | 0.15|    | 1.000         |
| Sweden        | 502  | 507  | 0.912   | 1.148 | 0.42        | 0.29|    | 1.000         |
| United Kingdom| 500  | 500  | 1.244   | 0.757 | 0.18        | 0.24|    | 0.021         |
| United States | 949  | 1006 | 1.125   | 1.001 | 0.36        | 0.33|    | 1.000         |

Note: Means and proportions weighted by sample-inclusion probability. Last two columns show null-hypothesis tests of the no design effect assumption proposed by Blair and Imai (2012: sec. 3.1). \( p_{BH} \) denotes q-values additionally adjusted for false-discovery rates of multiple-country comparisons using the Benjamini-Hochberg procedure (with \( \alpha = 0.05 \)).

substantive conclusions (note that only our macro plot in Figure I pools information from different countries).

B.4. Sample characteristics at baseline for treatment and control units

Table B.3 provides an overview of basic individual characteristics for respondents assigned to treatment and control groups. For each, the first column displays means followed by the standard error of the mean. The third column indicates the sample standard deviation. The final column lists the difference in means between treated and control groups. While randomization guarantees balance on covariates in expectation, we also find that observable characteristics in our sample are fairly balanced between treatment and control groups. Slightly more noticeable differences emerge for average ages in France, New Zealand, and Sweden, where members of the treatment group are about one to two years older. We do provide specifications that account for age differences in our estimates of individual-level determinants of non-compliance.
Table B.3
Covariates at baseline

| Country      | Control | Treated | Diff  |
|--------------|---------|---------|-------|
|              | Mean    | s.e.    | s.d.  | Mean    | s.e.    | s.d.  | (4-1) |
| Australia    | Age     | 45.33   | 0.63  | 14.36   | 45.29   | 0.65  | 14.47 | −0.040 |
|              | Female  | 0.49    | 0.02  | 0.50    | 0.49    | 0.02  | 0.50  | 0.001 |
|              | Ideology| 5.39    | 0.10  | 2.13    | 5.73    | 0.11  | 2.16  | 0.341 |
|              | Trust   | 0.53    | 0.02  | 0.50    | 0.50    | 0.02  | 0.50  | −0.028 |
| Austria      | Age     | 47.45   | 0.77  | 17.29   | 46.63   | 0.72  | 16.03 | −0.822 |
|              | Female  | 0.50    | 0.02  | 0.50    | 0.53    | 0.02  | 0.50  | 0.036 |
|              | Ideology| 5.03    | 0.09  | 2.04    | 4.93    | 0.09  | 2.05  | −0.098 |
|              | Trust   | 0.47    | 0.02  | 0.50    | 0.47    | 0.02  | 0.50  | −0.005 |
| Germany      | Age     | 48.49   | 0.52  | 16.57   | 49.30   | 0.52  | 16.56 | 0.803 |
|              | Female  | 0.53    | 0.02  | 0.50    | 0.49    | 0.02  | 0.50  | −0.038 |
|              | Ideology| 4.73    | 0.06  | 1.90    | 4.71    | 0.06  | 1.99  | −0.022 |
|              | Trust   | 0.48    | 0.02  | 0.50    | 0.50    | 0.02  | 0.50  | 0.024 |
| France       | Age     | 49.23   | 0.51  | 16.27   | 50.84   | 0.50  | 15.74 | 1.610 |
|              | Female  | 0.54    | 0.02  | 0.50    | 0.51    | 0.02  | 0.50  | −0.025 |
|              | Ideology| 5.42    | 0.08  | 2.34    | 5.20    | 0.08  | 2.35  | −0.022 |
|              | Trust   | 0.36    | 0.02  | 0.48    | 0.39    | 0.02  | 0.49  | 0.027 |
| United Kingdom| Age   | 46.54   | 0.72  | 16.09   | 46.56   | 0.73  | 16.37 | 0.019 |
|              | Female  | 0.50    | 0.02  | 0.50    | 0.51    | 0.02  | 0.50  | 0.033 |
|              | Ideology| 4.91    | 0.11  | 2.30    | 4.86    | 0.10  | 2.14  | −0.048 |
|              | Trust   | 0.34    | 0.02  | 0.47    | 0.36    | 0.02  | 0.48  | 0.022 |
| Italy        | Age     | 48.84   | 0.76  | 17.05   | 49.47   | 0.75  | 16.65 | 0.637 |
|              | Female  | 0.52    | 0.02  | 0.50    | 0.52    | 0.02  | 0.50  | 0.004 |
|              | Ideology| 4.85    | 0.11  | 2.37    | 5.20    | 0.12  | 2.60  | 0.354 |
|              | Trust   | 0.34    | 0.02  | 0.47    | 0.36    | 0.02  | 0.48  | 0.022 |
| New Zealand  | Age     | 45.34   | 0.72  | 16.14   | 47.72   | 0.78  | 17.42 | 2.376 |
|              | Female  | 0.52    | 0.02  | 0.50    | 0.52    | 0.02  | 0.50  | −0.005 |
|              | Ideology| 5.06    | 0.11  | 2.26    | 5.09    | 0.11  | 2.21  | 0.034 |
|              | Trust   | 0.58    | 0.02  | 0.49    | 0.59    | 0.02  | 0.49  | 0.012 |
| Sweden       | Age     | 48.06   | 0.72  | 16.18   | 49.15   | 0.73  | 16.43 | 1.097 |
|              | Female  | 0.53    | 0.02  | 0.50    | 0.48    | 0.02  | 0.50  | −0.044 |
|              | Ideology| 5.13    | 0.12  | 2.51    | 5.13    | 0.12  | 2.64  | 0.002 |
|              | Trust   | 0.58    | 0.02  | 0.49    | 0.54    | 0.02  | 0.50  | −0.033 |
| United States| Age    | 46.11   | 0.55  | 17.28   | 45.87   | 0.54  | 17.21 | −0.233 |
|              | Female  | 0.51    | 0.02  | 0.50    | 0.50    | 0.02  | 0.50  | −0.007 |
|              | Ideology| 5.70    | 0.09  | 2.75    | 5.38    | 0.09  | 2.93  | −0.329 |
|              | Trust   | 0.45    | 0.02  | 0.50    | 0.43    | 0.02  | 0.50  | −0.021 |
C. Latent variable model behavioral health changes

In this section we describe the construction of a one-dimensional latent factor capturing behavioral changes in response to the pandemic on a broader scope than our list experiment.

Our survey contains a battery of items asking respondents if they have changed their behavior since the beginning of the pandemic. These were placed distantly after the survey experiment. They are presented with a list of items:

- washing your hands more often and/or for a longer amount
- coughing or sneezing into your elbow or a tissue
- stopped greeting others by shaking hands, hugging or kissing
- keep a distance of [six feet] between yourself and other people\(^4\)
- reduced your trips outside home
- avoid busy places (public transportation, restaurants, sport)
- stopped seeing friends

Responses for each item were originally recorded on an 11-point response scale, ranging from 0 (“Not at all”) to 10 (“Yes, completely”). This question format produces extreme skewness of responses: for many items more than 50% of respondents chose the highest two out of 11 categories. We dichotomized all items such that responses other than ‘9’ and ‘10’ (the highest two categories) indicate that respondents likely did not adjust their behavior (or only did so in a selective manner).

We first study if the configuration of these items follows similar patterns in each country and if they can be summarized by a low-dimensional vector of latent variables. Figure C.1 summarized results from a nonlinear principal components analysis (Gifi 1990) of our dichotomized item battery estimated separately in each country. Panel A shows the eigenvalues for seven principles components in each country. It suggest that one component captures a dominant share of variation in each country. All eigenvalues for components other then the first are less then 1 (save for Sweden, which is barely above 1 for component 2). Similarly, panel B, which plots component loadings for each item on the first two principal components for each country, suggests that the predominant variation takes place on the first component. Based on this initial exploratory analysis, we specify one-dimensional latent factor/IRT models described next.

A simple latent variable model for these items would be a standard two-parameter IRT model estimated on the pooled sample. The parameters of this model are item intercepts, \(\tau\), (referred to as “difficulties” in the IRT literature) and coefficients, \(\lambda\), capturing how a unit increase in the latent variable shifts the propensity of observing each item (“discrimination

\(^4\)The distance used in this item corresponds to the health guidelines of each country at the time: 6 feet in New Zealand, UK, US; 3 feet in Australia, 1m in Austria, France, Italy; 2m in Germany; 1.5m in Sweden.
parameters”). Expressed briefly, and using factor-analytic notation (Takane and de Leeuw 1987), for any given item \( y \) the model takes the form \( y_i = \tau + \lambda f_i + \epsilon_i \), where the distribution of residuals \( \epsilon \) is normal with variance fixed to 1 and the latent variable \( f \) is distributed normally with mean 0 and unit standard deviation (for a detailed introduction to IRT models see, e.g., van der Linden 2016; Hambleton et al. 1991). Estimating this model in a Bayesian framework using the Gibbs sampler, draws from the posterior distribution of \( f \) can be obtained straightforwardly and aggregated to country-specific averages.

![Figure C.1](image_url)

**Figure C.1**

Nonlinear principal components analysis of behavioral adjustment battery.

Panel A shows the eigenvalues of seven principal components for each of 9 countries. It indicates that extracting one component captures a large proportion of total variation. Panel B shows component loadings plot for the first two largest components in each country. The configuration of the loadings in each country also suggests that a one-dimensional factor captures the most important differences between respondents.
However, the pooled model ignores the problem of measurement equivalence. Pooling information from different countries with potentially heterogenous response processes might make it invalid to compare means of the latent factor (see, e.g., Davidov et al. 2014; Stegmueller 2011). The factor analytic literature usually distinguishes between different degrees of measurement invariance (e.g. Millsap 2011): configural invariance assumes a similar fundamental factor structure in each country (as emerged in our PRINCALS analysis above), but puts no equality restrictions on any model parameters in different countries. Metric invariance adds equality constraints for loadings, while scalar invariance adds equality constrains for both loadings and intercepts. Under essential country heterogeneity in response processes, factor means and variances are only identified under the scalar invariance restriction. Thus, imposing equality in loadings and intercepts in the pooled model where it does not exist leads to distorted estimates of the latent factor and the resulting country means are not quantitatively comparable.

C.1. Random coefficient hierarchical factor model

Our preferred model is a latent variable model that explicitly allows for country-differences in differential item functioning following the proposals by de Jong et al. (2007) and Fox and Verhagen (2010). They key idea is to specify a hierarchical factor model with random coefficients (Ansari et al. 2000, 2002) allowing for heterogeneity in item parameters while being anchored to a common mean.

Denote by $y_{ijk}$ the response of person $i$ ($i = 1, \ldots, N_j$) in country $j$ ($j = 1, \ldots, J = 9$) to survey item $k$ ($k = 1, \ldots, K = 7$) probing if they changed health-relevant behaviors. Each item is specified as a probit equation and we work with the underlying latent variables $y_{ijk}^*$, which are available via data augmentation during the Gibbs sampler (Albert and Chib 1993). We specify each $y_{ijk}^*$ as being driven by an underlying latent factor $\theta_W$. We estimate the following measurement system

\[
y_{ij1} = \tau_{j1} + \lambda_{j1} \theta_{Wij} + \epsilon_{ij1} \\
\vdots \\
y_{ijk} = \tau_{jk} + \lambda_{jk} \theta_{Wij} + \epsilon_{ijk}
\]  

(C.1)

where $\tau$ are item intercepts, $\lambda$ are factor loadings and $\theta_{Wij}$ is a latent factor representing individual propensity to change behavior. For identification, $\theta_{Wij} \sim N(0, 1)$ as in standard IRT models.\(^5\) Residuals $\epsilon_{ijk}$ are also called uniquenesses in the factor analysis literature and are assumed independent after conditioning on the latent trait and distributed mean zero with unit variance (in order to fix the underlying variance of the probit model). Both item

\(^5\)The sign of the latent variable is not identified (Anderson and Rubin 1956). In our application this is of no concern since its orientation (“less” inclined to follow health guidance) is easily established from the pattern of loadings.
intercepts and loadings are free to vary over countries and are anchored by the following hierarchical factor structure:

\[
\begin{align*}
\tau_{jk} &= \tau_k + \lambda \theta_B j + \zeta_{\tau_{jk}} \quad (C.2) \\
\lambda_{jk} &= \lambda_k + \lambda \theta_{\psi_j} + \zeta_{\lambda_{jk}} \quad (C.3) \\
\theta_B &\sim N(0, \sigma^2_B) \quad (C.4) \\
\theta_{\psi j} &\sim N(0, \sigma^2_{\psi}) \quad (C.5)
\end{align*}
\]

where random item effects are distributed \( \zeta_{\tau_{jk}} \sim N(0, \sigma^2_{\tau}) \) and \( \zeta_{\lambda_{jk}} \sim N(0, \sigma^2_{\lambda}) \).

Fox (2010) discusses the identification constraints needed to separately identify varying factor means and variances with both intercepts and loadings hierarchically modeled. We follow the strategy outlined in Asparouhov and Muthén (2015). Note that the loadings \( \lambda \) are equal for \( \theta_B j \) and \( \theta_{\psi j} \).

The variation in item intercepts over countries is captured by \( \sigma^2_{\tau} \) while the variation in loadings is captured by \( \sigma^2_{\lambda} \). The systematic country-variation of individual factor means is captured by \( \sigma^2_B \); the variation in the factor variance is captured by \( \sigma^2_{\psi} \).

The model is estimated using Gibbs sampling using latent data augmentation for the dichotomous variables. We specify normal priors for all \( \lambda \) and \( \tau \) with mean 0 and prior variance 10. Random effect variance terms are given inverse Gamma priors with shape and scale set to 0.001. The prior for covariance matrix of the two factor variances is inverse Wishart with \( V = \text{diag}(2) \) and degrees of freedom set to \( \nu = \text{dim}(V) + 1 = 3 \).

C.2. Estimates and comparison to list experiment

Table C.1 shows estimates for all model parameters. Columns \( \tau \) and \( \lambda \) represent means of the threshold and loading random coefficients. Displayed are estimates (posterior means) with 95% credible intervals in brackets. The \( \sigma^2_{\tau} \) and \( \sigma^2_{\lambda} \) columns display the estimated variances of the random coefficients. We find that all seven items are strongly and significantly related to the latent factor. The two items likely relating to personal contact in public—“shaking hands or hugging” and “avoiding busy places”—show the highest discrimination parameter estimates and have very high estimated thresholds/difficulties as well. The private behavior of ‘washing hands often’ discriminates somewhat less.

Figure C.2 plots the resulting distribution of latent factors in nine countries. Panel A plots the “overall” latent variable, which combines both within and between country components estimated separately in the RC-IRT model. As expected, one can discern systematic country-specific mean shifts in the health-behavior latent factor. However, as panel B shows, once we focus on the within-country latent variable, \( \theta_W \), we are left with a latent variable for
Table C.1
Bayesian Hierarchical IRT model of propensity to not follow health guidelines.

| Ignored guideline                  | $\tau_k$ | $\sigma^2_{\tau_k}$ | $\lambda_k$ | $\sigma^2_{\lambda_k}$ |
|------------------------------------|----------|----------------------|-------------|------------------------|
| Washing hands often                | 0.387    | 0.194                | 0.952       | 0.07                   |
|                                    | [-0.04, 0.801] | [0.667, 1.261]       |             |                        |
| Cough/sneeze into tissue           | 0.327    | 0.254                | 0.892       | 0.03                   |
|                                    | [-0.099, 0.776] | [0.674, 1.142]       |             |                        |
| No shaking hands, hugs             | 1.661    | 0.116                | 1.669       | 0.016                  |
|                                    | [1.096, 2.26] | [1.289, 2.068]       |             |                        |
| Keep six feet distance             | 0.856    | 0.031                | 1.458       | 0.013                  |
|                                    | [0.381, 1.336] | [1.139, 1.815]       |             |                        |
| Reduce trips outside home          | 0.547    | 0.481                | 1.257       | 0.088                  |
|                                    | [-0.041, 1.187] | [0.921, 1.601]       |             |                        |
| Avoid busy places                  | 1.083    | 0.068                | 1.565       | 0.023                  |
|                                    | [0.555, 1.602] | [1.211, 1.949]       |             |                        |
| Stopped seeing friends             | 0.717    | 0.259                | 1.152       | 0.012                  |
|                                    | [0.213, 1.221] | [0.892, 1.43]        |             |                        |
| $\sigma^2_B$                       | 0.153    |                      |             |                        |
| $\sigma^2_u$                       | 0.272    |                      |             |                        |
| N                                  | 11,016   |                      |             |                        |

Note: Based on 10,000 MCMC samples. Using factor-analytic probit parametrization of 2-parameter IRT model; entries show random coefficients for intercept/difficulty parameters, $\tau$, and loadings/discrimination parameters, $\lambda$. Columns $\sigma^2$ show corresponding variances of country-specific deviations. Prior on latent factor variances is $IW(I_2, df = 3)$; priors on item parameters are $\pi(\tau) \sim N_K(0, sI_K)$ and $\pi(\tau) \sim N_K(0, sI_K)$ with $s = 100$; on item variance components $\forall k: \pi(\sigma^2_{\tau_k}) \sim IG(a_0, b_0), \pi(\sigma^2_{\lambda_k}) \sim IG(a_0, b_0)$ with $a_0 = b_0 = 0.001$.

individual differences in health-related behavior that is broadly comparable across different countries in terms of location and scale and shape.

Finally, we compare estimates of our latent variable of health behavior to estimates of non-compliance with social distancing guidelines from our list experiment. Table C.2 shows country-aggregates of latent factor estimates for each country compared to list experimental estimates of non-compliance with social distancing. Note that both values are not directly comparable. First, the list experiment is focused on one behavior—social distancing—, while the latent factor measures a broader tendency to change health-related behaviors. Second, the different scaling of both quantities makes numerical comparisons difficult: the latent variable is normalized to have mean 0 (with a fixed standard deviation of 1) while estimates from the list experiment lie in the unit interval. Nonetheless, comparing the rank order of estimates reveals that estimates from the list experiment follow a pattern comparable to estimates from the latent variable model. Sweden, Australia, and Germany show the largest factor estimates and are also among the four largest experimental estimates (with the exception of Austria). The United Kingdom and Italy, both with list experimental estimates of essentially zero also
Latent components of non-compliance with health guidelines

This figure plots kernel densities of latent variable distributions for 8 countries. The sum of all components (a) includes systematic differences in response behavior by populations of different countries. This corresponds to what one would obtain from a simple pooled IRT model. In contrast, the latent variable $\theta_W$ is the within-component net of country differences in item intercepts and factor loadings (b). We use the latter in our analyses below. Kernel density estimates (Gaussian kernel with bandwidth 0.4 evaluated over a 200-point grid) of 1,000 MCMC draws from posterior distribution of latent variables.

...emerge among the bottom three countries ranked via the latent factor estimates. The rank correlation between both sets of estimates is 0.73 with an exact $p$-value of 0.031.
Table C.2
Relationship between experimental estimates and country values of one-dimensional latent factor of health guidance behavioral adjustments.

| Country          | List experiment | Latent factor RC-IRT |
|------------------|-----------------|----------------------|
|                  | Mean difference |                      |
| A. Estimates     |                 |                      |
| Germany          | 0.640           | 0.308                |
| Sweden           | 0.484           | 0.464                |
| Austria          | 0.425           | −0.071               |
| Australia        | 0.336           | 0.330                |
| United States    | 0.209           | 0.236                |
| France           | 0.125           | −0.452               |
| New Zealand      | 0.120           | 0.052                |
| Italy            | 0.007           | −0.469               |
| United Kingdom   | −0.024          | −0.196               |
| B. Rank correlation with list experiment | | |
| Spearman’s ρ     | 0.73            |                      |
| Exact p-value    | 0.031           |                      |
| Parameters       | 31              |                      |
| N                | 12,028          |                      |
D. Nonparametric estimates details

D.1. Model and estimates

As discussed in section 4.2 in the main text, the model we estimate is:

\[
\theta_{W,ir} = \beta_0 + x_i^\prime \beta + f(z_i) + \xi_r + \epsilon_{ir},
\]

with the flexible function of ideology approximated via

\[
f(z) = \sum_{l=1}^{L} \gamma_l B_l(z).
\]

The B-spline coefficients \(\gamma\) are penalized using a quadratic penalty via a smoothing parameter \(\lambda\):

\[
\lambda^* \gamma^\prime K \gamma,
\]

where the positive (semi-)definite precision matrix \(K = D_r^\prime D_r\) and \(D_r\) is a matrix of \(r\)-th order differences (Eilers and Marx 1996). It is created by applying a \(r\)-th order difference operator \(\Delta^r\) to the spline coefficients, e.g., \(\Delta^1 = \gamma_l - \gamma_{l-1}\) for \(r = 1\). An example of a first order difference matrix, \(D_1\), and the resulting penalty matrix, for a setting with only 5 spline coefficients (for reasons of space) is shown below.

\[
D_1 = \begin{bmatrix}
-1 & 1 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & -1 & 1 \\
\end{bmatrix}, \quad K = D_1^\prime D_1 = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 \\
-1 & 2 & -1 & 0 & 0 \\
0 & -1 & 2 & -1 & 0 \\
0 & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & -1 & 1 \\
\end{bmatrix}
\]

As is common in the applied literature, in our application we employ second-order difference penalties. One can think of this as the Bayesian stochastic analog to the well known penalized regression approaches in a classical setting. The penalty expressed in the form of a Gaussian prior on \(\gamma\) is then given by:

\[
p(\gamma|\omega^2) \propto \left(\frac{1}{\omega^2}\right)^{\frac{r_0(K)}{2}} \exp\left(-\frac{1}{2\omega^2} \gamma^\prime K \gamma\right)
\]

and \(\lambda^* = 1/\omega^2\).

The “smoothness” of the estimated function is thus influenced directly by \(\omega^2\). In a Bayesian framework, we learn about \(\omega^2\) by assigning it a prior distribution. A “default” prior choice might be an inverse gamma distribution with small values for shape and rate. However, the
parametrization of this distribution is difficult to link back to the smoothing behavior of the prior in (D.4).

Simpson et al. (2017) propose a strategy to select priors for complexity penalties in a principled way. We briefly summarize their key principles here (for more details, see Simpson et al. 2017: 7–8):.

- Parsimony: prefer simple over complex unless the data suggest otherwise. The prior alone should prefer a simple model over a complex one. It thus should decay as a function of some measure of increasing complexity

- Complexity: captured by the distance \(d(p||p_0)\) between a more complex and a baseline model \(p_0\):
  \[
  d(p||p_0) = \sqrt{2 \times KLD(p||p_0)}
  \]  
  (D.5)
  where \(KLD\) is the Kullback-Leibler divergence \(KLD(p||p_0) = \int p(u) \log \left( \frac{p(u)}{p_0(u)} \right) du\)

- Constant rate penalty: penalize the deviation from a simple model parametrized via distance \(d\) with a constant rate of decay between simple and complex. This implies an exponential prior on the distance scale such that the mode of the prior is the simple (linear) model: \(p(d) = r \exp(-rd)\). The speed of decay, \(r\) is a hyperparameter to be set by the researcher.

Klein et al. (2016) show that implement these principles implies a Weibull prior with shape \(a = 1/2\) and scale \(\nu\) specified by the rate of decay \(r\) defined above (cf. Klein et al. 2016 Theorem 1 and appendix A.1)

\[
p(\omega^2) = \frac{1}{2\nu} \left( \frac{\omega^2}{\nu} \right)^{-\frac{1}{2}} \exp \left[ -\left( \frac{\omega^2}{\nu} \right)^{\frac{1}{2}} \right]
\]  
(D.6)

We use this prior in our specifications used in the main text, with \(\nu\) set to 0.01 (cf. Klein et al. 2016: Appendix B, esp. Table B1 and Figure B1). As Figure D.1 below shows, using alternative variance prior choices, such as the half-Cauchy prior on \(\omega\) (Gelman 2006) or half-Normal prior yields ideology function estimates that are visually almost indistinguishable. Note that the venerable “default” prior for variances, the inverse-Gamma prior, would put 0 weight on \(p_0\) (the simple linear model).

The model is completed by assigning priors to the remaining parameters. These are more straightforward and strongly dominated by the data. For regression-type coefficients we assign normal priors with large variance \(p(\beta) \sim N(0, 1E6)\). The prior for the variance of the random effects is \(\sigma_x^2 \sim G^{-1}(a_0, b_0)\).

Table D.1 shows central model parameters (we do not display individual spline coefficients for reasons of space) for two central specifications. Specification A is the one used in the main text. Specification D adds an indicator variable equal to one if a respondent cast a
vote in support of the government in the last election, while specification E uses a measure of how much respondents trust the head of the executive (the scale of the coefficients is in standard deviations). The corresponding parameter estimates (posterior means with posterior standard deviations in parentheses) for the latter two specifications are given by the rows labelled $\delta$. In many countries, we find rather strong smoothing indicating that estimates of the functional form of ideology will be close to a simple linear model. However, the United States and Italy and, to some degree, Austria and Australia are notable exceptions. To give a sense of scale, the penalty term in Germany is seven times larger than in the US and more than 4.5 times larger than in Austria.
### Table D.1
Parameter estimates for semi-parametric models of ideology and health-behavior

| Specification A | AUS | AUT | FRA | DEU | ITA | NZL | SWE | UK | USA |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| dim(γ)          | 9   | 9   | 9   | 9   | 9   | 9   | 9   | 9   | 9   |
| EDF             | 2.82| 3.01| 2.12| 2.29| 3.86| 2.36| 2.37| 2.98| 4.41|
| \(\omega^2\)   | 0.0026| 0.0031| 0.0003| 0.0007| 0.0048| 0.0012| 0.0007| 0.0020| 0.0051|
| \(\sigma^2_\xi\)| 0.0045| 0.0037| 0.0024| 0.0027| 0.0055| 0.0057| 0.0110| 0.0042| 0.0037|
| LogLik          | -1062| -1164| -2178| -2377| -1081| -1026| -1143| -1038| -2313|
| N               | 847 | 957 | 1825| 1923| 891 | 835 | 958 | 881 | 1883 |

| Specification D | δ   |       |       |       |       |       |      |      |      |
|-----------------|-----|-------|-------|-------|-------|-------|------|------|------|
|                 |     | 0.217 | -0.082| -0.025| -0.168| -0.047| -0.015| -0.021| -0.086| 0.119|
|                 |     | (0.066)| (0.054)| (0.057)| (0.041)| (0.060)| (0.063)| (0.066)| (0.067)| (0.048)|
| dim(γ)          | 9   | 9     | 9     | 9     | 9     | 9     | 9    | 9    | 9    |
| EDF             | 3.14| 3.07  | 2.25  | 2.34  | 3.84  | 2.27  | 2.38 | 2.91 | 4.43 |
| \(\omega^2\)   | 0.0035| 0.0032| 0.0006| 0.0008| 0.0046| 0.0010| 0.0007| 0.0018| 0.0051|
| \(\sigma^2_\xi\)| 0.0049| 0.0036| 0.0025| 0.0026| 0.0052| 0.0055| 0.0113| 0.0043| 0.0036|
| LogLik          | -1056| -1163| -1596| -2369| -1081| -1025| -1143| -1038| -2310|
| N               | 847 | 957 | 1395 | 1923 | 890 | 834 | 958 | 881 | 1883 |

| Specification E | δ   |       |       |       |       |       |      |      |      |
|-----------------|-----|-------|-------|-------|-------|-------|------|------|------|
|                 |     | -0.088| -0.157| 0.003 | -0.100| -0.034| -0.156| -0.022| -0.075| 0.058^a|
|                 |     | (0.035)| (0.030)| (0.021)| (0.020)| (0.032)| (0.038)| (0.033)| (0.034)| (0.022)|
| dim(γ)          | 9   | 9     | 9     | 9     | 9     | 9     | 9    | 9    | 9    |
| EDF             | 2.96| 3.09  | 2.16  | 2.37  | 3.85  | 2.29  | 2.35 | 3.03 | 4.44 |
| \(\omega^2\)   | 0.0029| 0.0037| 0.0004| 0.0008| 0.0047| 0.0010| 0.0007| 0.0020| 0.0053|
| \(\sigma^2_\xi\)| 0.0044| 0.0035| 0.0023| 0.0027| 0.0059| 0.0056| 0.0111| 0.0044| 0.0034|
| LogLik          | -1058| -1151| -2177| -2365| -1077| -1018| -1143| -1036| -2310|
| N               | 846 | 957 | 1824 | 1923 | 890 | 835 | 958 | 881 | 1883 |

Note: \(\delta\) refers to the coefficient for government vote and trust in head of executive variables (specifications D and E, respectively). Shown are posterior means with posterior standard deviation in parentheses. \(\text{dim}(\gamma)\) indicates the number of spline coefficients, EDF is the estimated (posterior mean) effective degrees of freedom. \(\omega\) is the posterior mean of the inverse smoothing parameter in the spline coefficient prior. \(\sigma^2_\xi\) denotes the posterior mean of the variance parameter of the regional random effects. Other covariates, as well as individual spline coefficients, omitted to save space.

^a Data for the US lacks information on this variable and the model employs responses from the prior longitudinal wave and imputes missing responses in the current wave from the responses to related items observed in both survey waves using the random forest algorithm proposed by Stekhoven and Bühlmann (2012).
D.2. Accounting for government support

Panel A of Figure D.2 plots estimated ideology functions from specification A used in the main text (red line) and specifications D and E, which include an indicator capturing if a respondent voted for the current government in the last election, and a variable capturing respondents’ trust in the head of the executive, respectively. We selected the four countries where the differences in ideology function estimates were more noticeable. The inclusion of variables capturing support of the government (in a broad sense) does shift the ‘ideology curve’ somewhat, especially for respondents to the left of the scale. It shifts it downwards in Australia and the UK, and upwards in the US and Austria. However, the magnitude of this change is relatively minor.

![Figure D.2](image)

Impact of ideology when adjusting for government support

Panel A plots expected values of $\theta_W$ from our main model (A) and specifications adjusting for past pro-government vote (D) and trust in the head of the executive (E). Panel B plots differences with uncertainty represented by differences of 1,000 randomly drawn function evaluations.

Panel B plots the differences between specifications A and D and A and E. Uncertainty is represented by hairlines plotting 1,000 evaluated differences at 1,000 random draws from the posterior distribution. It underscores that the relative difference between these three specifications is rather small with a range of uncertainty always including zero.
D.3. Accounting for trust in experts

We did not include a measure of trust in (health) expertise in our main analysis, partly because such trust itself might be shaped by ideological predispositions. Indeed, in most countries under study (with the notable exception of the UK) trust in scientist shows non-zero correlations with ideology (where more right-leaning respondents trust scientists less).

However, it is germane to ask if our displayed relationship between ideology and health behavior mostly reflect individual differences in trust of experts, such as scientists and doctors. To explore this question, we employ indicators capturing how much a respondent trusts scientists and doctors. Both are measured on a four point scale with response options ranging from “don’t trust at all” to “trust completely”. In our analysis, we scale both variables to standard deviation units.

Figure D.3

Estimates of respondents’ trust in experts on health behavior

This figure plots changes in expected values of $\theta_W$ for a standard deviation increase of respondents’ trust in doctors (left panel) or scientists (right panel). Posterior means and 90% credible intervals based on 10,000 MCMC samples.

Figure D.3 shows the relationship between the two trust measures and respondents’ propensity to ignore health advice. It displays changes in expected values for a unit change in each variable. Figure D.3 indicates quite clearly that higher trust in scientists and doctors increases compliance with health advice. Notable exceptions are Sweden and France, where the respective estimates for trust in doctors and scientists are close to zero. It is instructive to compare estimates for the United States with those obtained in analyses of government support (reported in the main text). While the US was the only country where increased support of the government lead to higher propensity to ignore health guidelines, patterns for trust in doctors and scientists are in line with those found in other developed countries (this is especially true for trust in doctors, which is likely the less politically charged item). Given
the impact of trust in experts, we now investigate if and how this impacts our estimated ideology-health advice patterns.

Panel A of Figure D.4 plots estimated ideology functions from specification A used in the main text (red line) and specification E, which includes the two trust in experts variables. Panel B shows the difference between A and E with uncertainty represented by hairlines plotting 1,000 evaluated differences at 1,000 random draws from the posterior distribution. We focus on the three countries were differences were most pronounced. However, Figure D.4 reveals that accounting for individual differences in trust in experts has only limited impact on the estimated functional form of the relationship between ideology and the propensity to not follow health advice.
D.4. Model comparisons

To compare complex model specifications we need an information criterion that penalizes model deviance for the number of parameters used. In models with heavy regularization (like the ones in this paper), the effective number of parameters can be much lower than the numbers of parameters one would obtain by counting coefficients. In the main text, we use the WAIC (Watanabe 2013), which penalizes the Bayesian model deviance by the number of effective model parameters. We estimate the number of effective parameters using the variance-based calculation proposed in Gelman et al. (2013: 173).

We prefer WAIC over the more widely used DIC due to its somewhat more attractive properties. First, one criticism of the DIC is that it uses a point estimate (Van Der Linde 2005; Plummer 2008), while WAIC uses the entire posterior parameter distributions. Second, WAIC is asymptotically equal to Bayesian leave-one-out cross-validation (Watanabe 2010). For an excellent discussion, see Gelman et al. (2014). However, our core findings regarding model comparisons do not depend on this choice. Table D.2 below shows model comparisons carried out using the DIC and its effective number of parameters.

| Specification          | AU  | AT  | FR  | DE  | IT  | NZ  | SE  | UK  | US  |
|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A: Basic demographics  | 2552| 2435| 4859| 4964| 2463| 2428| 2431| 2408| 4964|
|                        | 7.9 | 8.3 | 11.2| 9.7 | 10.0| 8.6 | 10.1| 9.5 | 11.9|
| B: f(Ideology)         | 2134| 2340| 4369| 4766| 2176| 2063| 2298| 2089| 4644|
|                        | 11.1| 12.0| 12.4| 12.3| 14.2| 11.7| 13.0| 12.6| 17.2|
| C: Region differences  | 2135| 2340| 4370| 4761| 2176| 2063| 2299| 2090| 4644|
|                        | 12.1| 15.2| 17.9| 27.3| 16.0| 14.6| 14.7| 14.1| 20.9|
| D: Government vote     | 2125| 2339| 3205| 4752| 2177| 2062| 2300| 2090| 4639|
|                        | 12.6| 12.9| 13.3| 13.4| 14.9| 12.5| 14.1| 13.6| 18.2|
| E: Trust in executive  | 2128| 2315| 4367| 4743| 2169| 2049| 2300| 2086| 4638|
|                        | 12.2| 13.1| 13.3| 13.4| 15.3| 12.7| 14.0| 13.7| 17.8|

*Note: DIC and effective number of parameters estimated from 10,000 MCMC samples.*

D.5. Comparisons to pooled linear fixed effects models

Figure D.5 illustrates the difference between semiparametrically estimated ideology-health functions and the relationship one would obtain when specifying a “typical” linear fixed effects model. Estimating this model (that is, regressing the latent factor on z-standardized ideology, a set of controls, and country fixed effects) yields a coefficient on ideology of 0.046 with a (cluster robust) standard error of 0.011. This would indicate a substantive and statistically significant effect of ideology on average. However, while this average fits
well for Germany and, to some degree, New Zealand, it is misleading for countries such as France (where the relationship is essentially zero) and Italy or the United States (where the relationship is strongly nonlinear), illustrating once more the risk of “pooling disparate observations” (Bartels 1996).

![Figure D.5](image)

**Figure D.5**

**Comparison of semiparametric and linear fixed effects models**

This figure illustrates the difference between semiparametric estimates of ideology and a typical fixed effects linear model specification, which regresses $\theta_w$ on ideology, a set of controls and country fixed effects.
References

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422), 669–679.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Volume 5, Berkeley, pp. 111–150. University of California Press.

Ansari, A., K. Jedidi, and L. Dube (2002). Heterogeneous Factor Analysis Models: A Bayesian Approach. *Psychometrika* 67(2), 49–78.

Ansari, A., K. Jedidi, and S. Jagpal (2000). A hierarchical bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science* 19(4), 328–347.

Asparouhov, T. and B. Muthén (2015). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications*. Charlotte, NC: Information Age Publishing.

Bartels, L. M. (1996). Pooling disparate observations. *American Journal of Political Science* 40(3), 905–942.

Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis* 20(1), 47–77.

Davidov, E., B. Meuleman, J. Cieciuch, P. Schmidt, and J. Billiet (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology* 40, 55–75.

de Jong, M., J.-B. Steenkamp, and J.-P. Fox (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical irt model. *Journal of Consumer Research* 34, 260–278.

Dong, E., H. Du, and L. Garden (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet* 20(5), P533–P534.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-Splines and Penalties. *Statistical Science* 11(2), 89–121.

Fox, J.-P. (2010). *Bayesian Item Response Modeling. Theory and Applications*. New York: Springer.

Fox, J.-P. and A. J. Verhagen (2010). Random item effects modeling for cross-national survey data. In E. Davidov, P. Schmidt, and J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications*, pp. 467–488. London: Routledge.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1(3), 515–534.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis*. London: CRC Press.
Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing* 24(6), 997–1016.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester, NY: Wiley.

Hale, Hale, T., S. Webster, A. Petherick, T. Phillips, and B. Kira (2020). Oxford COVID-19 government response tracker,. Technical report, Blavatnik School of Government.

Hambleton, R. K., H. Swaminathan, and H. J. Rogers (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage.

Klein, N., T. Kneib, et al. (2016). Scale-dependent priors for variance parameters in structured additive distributional regression. *Bayesian Analysis* 11(4), 1071–1106.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor & Francis.

OECD (2020). *How’s Life? 2020: Measuring Well-being*. Paris: OECD Publishing.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics* 9(3), 523–539.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* 32(1), 1–28.

Stegmueller, D. (2011). Apples and oranges? the problem of equivalence in comparative research. *Political Analysis* 19, 471–487.

Stekhoven, D. J. and P. Bühlmann (2012). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1), 112–118.

Takane, Y. and J. de Leeuw (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 52(3), 393–408.

Van Der Linde, A. (2005). Dic in variable selection. *Statistica Neerlandica* 59(1), 45–56.

van der Linden, W. J. (2016). Unidimensional logistic response models. In W. J. van der Linden (Ed.), *Handbook of Item Response Theory. Volume I. Models*, pp. 13–30. CRC press.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(12), 3571–3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research* 14(Mar), 867–897.