Ideology and compliance with health guidelines during the COVID-19 pandemic: A comparative perspective.*

Michael Becher† Daniel Stegmueller‡ Sylvain Brouard§ Eric Kerrouche‡

Abstract

Objective: We measure the prevalence of non-compliance with public health guidelines in the COVID-19 pandemic and examine how it is shaped by political ideology across countries. *Methods:* A list experiment of non-compliance and a multi-item scale of health-related behaviors were embedded in a comparative survey of 11,000 respondents in nine OCED countries. We conduct a statistical analyses of the list experiment capturing degrees of non-compliance with social distancing rules and estimate ideological effect heterogeneity. A semi-parametric analysis examines the functional form of the relationship between ideology and the propensity to violate public health guidelines. *Results:* Our analyses reveal substantial heterogeneity between countries. Ideology plays an outsized role in the United States. No association of comparable magnitude is found in the majority of the other countries in our study. In many settings, the impact of ideology on health-related behaviors is non-linear. *Conclusion:* Our results highlight the importance of taking a comparative perspective. Extrapolating the role of ideology from the United States to other advanced industrialized societies might paint an erroneous picture of the scope of possible non-pharmaceutical interventions. Heterogeneity limits the extent to which policy-makers can learn from experiences across borders.

**Author contributions:* M.B., S.B. and D.S. conceived the study and designed the experiment. M.B., and D.S. analyzed the data and wrote the paper. E.K. provided additional funding. For comments and suggestions, we are grateful to Thomas Gschwend and participants of the conference "Public Policy, Opinions, Behavior, and Health Outcomes during the COVID Pandemic" organized by Lorena Barberia, Thomas Pluemper, Robert K. Goidel, and Guy Whitten. We also thank participants of (virtual) seminars at IAST, Berlin Social Science Center (WZB), and IE University. Stefan Preuß provided excellent research assistance. Becher acknowledges IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010. Brouard acknowledges the financial support from ANR – REPEAT grant (Special COVID-19), CNRS, Fondation de l'innovation politique, regions Nouvelle-Aquitaine and Occitanie. Stegmueller's research was supported by Duke University and the National Research Foundation of Korea (NRF-2017S1A3A2066657). The anonymous pre-analysis plan of the survey experiment is available at `http://aspredicted.org/blind.php?x=hv7yv2`. Ethics approval was obtained by The Review Board for Ethical Standards in Research at the Toulouse School of Economics and the Institute for Advanced Study (ref.code 2020-04-001).

†IE School of Global and Public Affairs, `michael.becher@ie.edu`
‡Duke University, `daniel.stegmueller@duke.edu`
§CEVIPOF, Science Po, Paris

1 Introduction

In the fight against epidemics, including the novel coronavirus disease (COVID-19) caused by the Sars-Cov2 virus, large-scale behavioral change is essential to limit the loss of human lives and to allow societies to resume economic and social activities. After the outbreak of COVID-19 in 2019 and its global spread as a pandemic in the first half of 2020, the absence of vaccination and medical treatment meant that non-pharmaceutical interventions—such as social distancing and hand washing—were crucial to mitigate and contain the spread of the virus. Most governments have adopted clear recommendations and rules to limit physical and social contact, and social scientists have immediately started to study individuals' compliance with these new behavioral rules. Surveys on COVID-19 in different countries have usually shown high rates of self-reported compliance with recommended health norms in the population (Barari et al. 2020; Brouard et al. 2020; Perrotta et al. 2020; Utych and Fowler 2020). At the same time, research, focused mostly on the United States (U.S.), has shown that political predispositions may undermine compliance with health guidelines (Allcott et al. 2020; Andersen 2020; Grossman et al. 2020; Kushner Gadarian et al. 2020; Painter and Qiu 2020). But we know less about whether ideology plays the same role in other countries.

In the face of high self-reported compliance, several scholars have cautioned that direct survey questions risk suffering from measurement error due to social desirability bias (Barari et al. 2020; Daoust et al. 2020).[1] While some have reported that online-mode surveys reduce the impact of desirability bias (Holbrook and Krosnick 2010), the public salience of these health measures may still induce overreporting of compliant behavior (Barari et al. 2020: 4; Munzert and Selb 2020). Social desirability is likely to be a factor when respondents are directly asked to report whether they complied with highly-publicized behavioral rules. We might not only underestimate the general rate of non-adherence to health guidelines, but also the importance of factors predicting it. Access to alternative measures of non-compliance, as well as credible estimates of sub-groups that are least likely to comply, is thus paramount to the design of effective non-pharmaceutical interventions during the pandemic (Bavel et al. 2020; West et al. 2020).

In this paper, we aim to make two contributions. First, we employ a list experiment to measure the *prevalence of non-compliance* with social distancing guidelines among the population of nine OECD countries.[2] Our sample covers approximately 65% of total confirmed COVID-19 related deaths at the time of the survey (Dong et al. 2020). It covers large variation in mortality (from less than five deaths per million inhabitants in Australia and New Zealand to more than 350 in Italy) and governmental responses (from very strong restrictions in Italy, New Zealand and France to comparatively few restrictions in Sweden). Our data reveals a substantial degree of non-adherence to social distancing guidelines. In most countries under study, experimental estimates of non-compliance are much higher than estimates based on (single item) direct questions from other surveys fielded in the same countries at the same time, which offer a more optimistic picture (Perrotta et al. 2020). Furthermore, we find large variation in rates of non-compliance across countries. This variation is not random, but varies systematically with the severity of the crisis and the stringency of lockdown-style policies.

Second, we examine factors driving *individual variation in non-compliance* with a particular focus on the role of political ideology, which has been identified as a salient marker for divergent health behavior during the pandemic.[3] Achieving compliance with collective decisions is a general problem that states tackle with a mix of monitoring, sanctions and voluntary cooperation. A large, cross-disciplinary literature on compliance suggests that for a given level of external enforcement, individuals may vary in their willingness to adapt their behavior to COVID-19 guidelines (Levi and Stoker 2000; Luttmer and Singhal 2014). A voluminous body of research in political science uses political ideology as a conceptual map structuring political beliefs and behavior (see Feldman 2013 for a recent overview). By shaping what information people seek out and how they interpret it, ideology can affect behavior beyond the voting booth (e.g., Campbell et al. 1960; Bartels 2002; Gerber and Huber 2009; Graham and Svolik 2020). The question of whether to comply or not with new health rules that require profound adjustments in one's daily life touches on the fundamental relationship between the state and its citizens. Different ideologies may entail different beliefs about health risks, the effectiveness of health measures or the legitimate scope of government action during the crisis (Barrios and Hochberg 2020; Kushner Gadarian et al. 2020). Relatedly, research in public health and the psychology of disease outbreaks has turned towards examining the role of (macro- and individual-level) social and psychological factors associated with health-related behaviors (e.g., Schaller 2011; Schaller and Park 2011). In both Europe and the U.S., systematic correlations between political ideology and health behavior and outcomes can be observed (Subramanian et al. 2009; Subramanian and Perkins 2009; Huijts et al. 2010). Our analysis thus moves beyond basic socio-demographic variables, such as age, gender, and education, which have received most attention in studies of compliance with non-pharmaceutical interventions, and examines the role of political ideology on the individual and cross-national level.

To conduct a comparative test of the association between ideology and compliance with health rules during the COVID-19 pandemic, we employ statistical tools for the analysis of heterogeneity in list experiments as well as a semi-parametric analysis of the impact of ideology on general pandemic health behavior. We measure the latter by estimating a cross-national item response theory model of a pandemic health behavior item battery (ranging from changes in hand-washing frequency to no longer meeting friends) included in our survey. Our results reveal substantial cross-national variation in the link between ideology and compliance. For the U.S., we find a strong association between ideology and compliance with health guidelines. Individuals that place themselves on the extreme right of the political spectrum are less likely to practice social distancing, measured using the experiment, than those with centrist views, whereas people with extreme left beliefs are more likely to comply with social distancing. Going beyond the list experiment and using a semiparametric latent variable model, we also find a clear ideological gap concerning broader behavioral adjustments in the U.S. These results are consistent with recent evidence from the U.S. on the partisan gap in elite rhetoric, perceived health risk and mass compliance between Democrats and Republicans (Allcott et al. 2020; Barrios and Hochberg 2020; Green et al. 2020; Grossman et al. 2020; Kushner Gadarian et al. 2020; Painter and Qiu 2020).

In contrast, we find that in most European countries under study political ideology is not strongly linked to compliance with social distancing in particular and health behavior adjustments more broadly during the initial stage of the pandemic. The absence of a clear association does not simply reflect a lack of variation in compliance, since it also emerges in countries with large non-compliance with social distancing (e.g., Austria and Germany). Rather, it may be due to the specifics of how the politics of the pandemic played out across countries. Thus, our results have implications for policy choices in response to the pandemic. The heterogeneity of individual-level results across countries suggests that it may be difficult to learn from other countries' experiences. Identifying the characteristics of non-compliers, who could be targeted or nudged into more compliance, might be a task that depends on country-specific political idiosyncrasies unlikely to be guided well by using results from other countries.

This article proceeds as follows. The following section describes the data, including the list experiment and pandemic health behavior item battery. Section 3 present experimental estimates of non-compliance rates and discusses identifying assumptions of the experiment. Section 4 examines the structure of heterogeneity in non-compliance due to individual- and country-level differences in political ideology. Section 5 provides an exploratory analysis of the link between country characteristics, such as the strictness of lockdown measures, and our health behavior estimates. Section 6 concludes.

2 Data

2.1 Survey data

Our list experiment and health behavior battery are embedded in a comparative internet survey covering eight countries in lockdown in mid-April 2020 (Australia, Austria, France, Germany, Italy, New Zealand, United States, United Kingdom) as well as Sweden. The latter took a less stringent policy response, but still recommended changes in health behavior. Surveys were conducted online between April 15-20, 2020 by established commercial polling companies (CSA Research in Australia and the U.S.; IPSOS in all other countries). All participants gave explicit consent to take part in the survey. Table A.1 in the appendix lists fieldwork periods, sample sizes, and the survey completion rate of participating respondents for each of the nine surveys. It also includes COVID-19 deaths and the overall strictness of lockdown-style measures. Non-completion is generally low. Target sample sizes were about 2,000 respondents in Germany, France, and the U.S., and 1,000 respondents in the remaining countries. Our final sample comprises 11,038 respondents. Sampling was done as part of existing online panels using quota sampling. The resulting samples were weighted by the survey providers to match Census population margins for gender, age, occupation, region, and degree of urbanization. All our analyses and descriptive results use probability weights unless otherwise indicated.

Our central variable, a respondent's ideology, is captured using a standard left-right or liberal-conservative 11-point self-placement item (the question reads: "on a scale from 0 to 10, where 0 is left and 10 is right, where would you place yourself politically?"; in the U.S., the wording is "liberal" and "conservative" rather than "left" and "right"). This left-right dimension tends to be the most salient dimension for most of the population in the countries under study (Huber and Inglehart 1995). It represents a simple one-dimensional conceptualization of political ideology dating back to the work of Converse on mass beliefs systems (Converse 1964). Jost et al. (2003) provide a recent theoretical justification for a uni-dimensional approach to political ideology.[4]

2.2 Non-compliance experiment

The list experiment or unmatched or item count technique (Miller 1984; Raghavarao and Federer 1979) allows respondents to truthfully report their behavior with respect to social distancing without revealing it to the researcher. Faced with a list of items, respondents are asked how many of these things they have done last week, but not which specific ones. Respondents are randomly assigned to treatment and control groups. As shown in Table I, the treated group received an additional item capturing the violation of the social distancing norm. The key assumption is that the treatment group would have responded like the control group absent the treatment. Our set of control items includes behaviors likely influenced by the pandemic but not violating health guidelines, such as ordering food using online delivery services, or exercising outdoors. The sensitive item presented only in the treatment group states that a respondent met with family or friends who are not part of the same household in the past week. This violates the societal norms in place during the pandemic.

Meeting friends and family not living in the same household ran counter the explicit government advice and stay-at-home orders in most countries under study during the time of the survey (Hale et al. 2020; Appendix Table B.1 summarizes social distancing rules in each country). Usually, people were only allowed to leave their home for "essential" trips such as work, health care, grocery shopping, daily exercise or emergencies. For instance, France banned all gatherings, and, in her first speech to the nation on the pandemic, German chancellor Angela Merkel told people to avoid visits to friends and family. In the U.S., stay-at-home orders were in place in 43 states, including the most populated ones. The main exception is Sweden, which explicitly did not prohibit private meetings with members of different households. However, while not imposing legal sanctions, it did appeal to people's personal responsibility.

Table I
List experiment items

List A: Control group	
1.	I went to the doctor or to the hospital
2.	I used public transportation to get to work
3.	I exercised outdoors
4.	I ordered food using an online delivery service
List B: Treatment group	
1.	I went to the doctor or to the hospital
2.	I used public transportation to get to work
3.	I exercised outdoors
4.	I met with two or more friends or relatives who do not live with me
5.	I ordered food using an online delivery service

Note: The list experiment is introduced by an identical statement for treated and control cases: "How many of these things have you done last week? You do not need to tell me which ones you have done, just how many."

2.3 Health behavior items

Our survey contains a battery of items capturing changes in health behavior following expert guidelines during the pandemic. The items were placed distantly after the survey experiment. Respondents were invited to indicate if they changed their behavior since the beginning of the pandemic with respect to a range of health-relevant actions listed in Table II below.

Table II
Health behavior adjustment items

Behavior	Prop. Unchanged[b]
Washing your hands more often and/or for a longer amount	0.395
Coughing or sneezing into your elbow or a tissue	0.399
Stopped greeting others by shaking hands, hugging or kissing	0.209
Keep a distance of [six feet] between yourself and other people[a]	0.326
Reduced your trips outside home	0.388
Avoid busy places (public transportation, restaurants, sport)	0.291
Stopped seeing friends	0.338

a Country-specific distancing guideline values are used
b Proportion of individuals indicating they did not change behavior. See appendix C for definition and details. Weighted by sample-inclusion probabilities.

A nonlinear principal component analysis (see Appendix Figure C.1) suggest that all seven items form a one-dimensional latent factor in each country under study. We conceptualize this latent variable as an individuals disposition of *health behavior adjustment* following expert public health guidelines. This fits with the widely-used definition of health behaviors, or health-related behaviors, as “actions taken by individuals that affect health or mortality [...] and can promote or detract from the health of the actor or others” (Short and Mollborn 2015). However, since there are stark differences in societal contexts (variation in expert guidelines, differences in pandemic severity, different social norms), we employ an item response theory (IRT) model that explicitly allows for cross-country variation in measurement parameters (de Jong et al. 2007; Fox and Verhagen 2010). Appendix C.2 provides more technical details, while Appendix C.3 provides parameter estimates and plots of the distribution of the country-specific latent factors. There we also show that *individual-level* estimates of the latent factor follow a similar rank order as *macro-level* rates of non-compliance with social distancing guidelines estimated from the list experiment: the Spearman rank correlation between both is 0.73 (with an exact p-value of 0.031).

3 Experimental estimates of non-compliance

In this section, we present estimates of the rate of non-compliance with social distancing guidelines based on our list experiment. One advantage of list experiments (Miller 1984; Raghavarao and Federer 1979) as a measurement device is that it reduces the risk of respondents’ answers being shaped by social desirability bias.

The key assumptions for identification in this design are (i) randomization of treatment (true in our survey by design), (ii) no design effects (i.e., responses to control items are not affected by the treatment), (iii) a truthful response to the sensitive item in the treatment condition under the anonymity awarded by the design (Imai 2011). We examined three possible empirical implications of violating assumptions (ii) and (iii) and generally find no evidence that the design is invalid. A first potential concern is that (anticipated) ceiling effects may undermine a the anonymity of a response: a respondent in the treatment group stating that she engaged in *all* of the listed acts would reveal her norm violation to the researcher and may thus not respond truthfully. Our set of questions deliberately uses innocuous control items unlikely to be all answered in the affirmative (or negative). Data from our experiment show that reported counts (in the control and treatment group) are not concentrated at the ceiling (see Table A.2). Furthermore, "self-administration" of the measurement instrument in an online survey context likely reduces non-truthful responses as well (Droitcour et al. 2011: 190). Second, to ensure to not to be associated with the sensitive item, the same individual who reports a non-zero count in the treatment group might want to counterfactually report a zero count in the treatment group. However, inspection of the data from our experiment shows that the share of respondents reporting zero counts is generally not higher under treatment than under control conditions. Third, we conducted statistical tests for the assumption of no design effects (Blair and Imai 2012) and do not reject the null hypothesis of no design effects (see Table A.2).

Table III
Experimental estimates of prevalence of individuals not following health guidelines during the COVID-19 pandemic in 9 countries.

Country	Prevalence	s.e.	95% CI	N
Australia	0.336	0.085	[0.17 : 0.50]	1007
Austria	0.425	0.082	[0.27 : 0.58]	996
France	0.125	0.041	[0.04 : 0.21]	2020
Germany	0.640	0.055	[0.53 : 0.75]	2000
Italy	0.007	0.067	[−0.12 : 0.14]	997
New Zealand	0.120	0.058	[0.01 : 0.23]	998
Sweden	0.484	0.077	[0.33 : 0.64]	1009
United Kingdom	−0.024	0.067	[−0.16 : 0.11]	1000
United States	0.196	0.071	[0.06 : 0.34]	1955

Note: Estimates based on difference-in-means between item count in the treatment and control group. Weighted by sample-inclusion probabilities.

Table III shows the estimated fraction of individuals in each country who met two or more friends or relatives not living in their household during the previous week. The estimates reveal a substantial degree of non-adherence to social distancing guidelines during the pandemic. In six out of eight countries a large and statistically significant fraction of the population did not follow social distancing guidelines. In some countries under lockdown (Austria and Germany), a (near-)majority of the population met friends or relatives against explicit recommendations. In the U.S. and Australia, a large minority (of at least 20% or more) of the population did not follow the norm. Experimental estimates of non-compliance are lower (but still statistically significantly different from zero) for France (13%) and New Zealand (12%). The fraction of non-compliers is not statistically distinguishable from zero in Italy and the U.K. Finally, in Sweden, which did not enact a lockdown and where social distancing rules were more permissive, around half of the population (48%) met friends or relatives. In the final empirical section, we explore whether variation in non-compliance across countries is systematically related to the severity of the pandemic and the strictness of lockdown measures.

4 Individual heterogeneity in non-compliance

We now turn to an analysis of heterogeneity in non-compliance focusing on the role of political ideology. The literature on compliance with government decisions points out that political beliefs may be relevant (Levi and Stoker 2000). If this is true in the case of COVID-19 as well, it indicates a considerable challenge for democratic governments trying to encourage compliance with non-pharmaceutical public health measures. While political beliefs are not easily changed and targeting interventions (e.g., messaging or surveillance) to different political groups raises important normative questions, it is important to understand whether political predispositions undermine compliance.

In this section, we employ two complementary empirical strategies. First, we will use the list experiment and estimate heterogeneity of non-compliance rates in terms of ideology (and other basic individual characteristics) using the estimator proposed by Imai (2011). The advantage of using the list experiment to examine sub-group heterogeneity is a clean identification of the model relying on the random assignment of respondents to treatment or control group and the use of pre-treatment covariates. One should keep in mind that we do not experimentally manipulate a respondent's ideology. Therefore, the resulting heterogeneity estimates describe a statistical association between ideology and the probability of complying with social distances and do not justify a causal interpretation based on the experimental design alone. In adjusted models we add pre-treatment covariates to reduce concerns about confounding. A potential downside of the list experiment is the limited power to detect smaller effects and the difficulty of estimating more flexible model specifications. To do the latter, we use individual-level estimates of the health behavior adjustment factor extracted using a random coefficient IRT model fit to our seven-item battery of health-related behavioral changes during the pandemic (see Appendix C.1). We fit a flexible semi-parametric regression model that allows us to examine in detail the structure of heterogeneity both within and among countries.

4.1 Heterogeneity in social distancing behavior

We fit beta-binomial regression models to the item count in each country (Imai 2011; Blair and Imai 2012) both including only the heterogeneity variable of interest and including an additional set of basic demographic controls (termed "unadjusted" and "adjusted" models). Missing observations are deleted listwise and the same set of observations are used in the unadjusted and adjusted specification. We conduct all analyses separately by country, in order to better isolate the contribution of individual-level factors from macro-level characteristics, such as variation in policies, institutions or the severity of the pandemic. Estimates are obtained using maximum likelihood using the Expectation-Maximization algorithm (Blair et al. 2020). Adjusted models reported in

Figure I include the following set of individual-level covariates: age (in years) and age squared; an indicator equal to 1 if female (0 otherwise); an indicator equal to 1 if a respondent has at least a college (BA) degree; subjective personal health measured on 5-point scale, political ideology (11-pt scale); an indicator for religiosity equal to 1 if a respondent feels close to any religion, and an indicator variable for interpersonal trust.[5] We measure ideology as discussed above and include it in both linear and quadratic form.

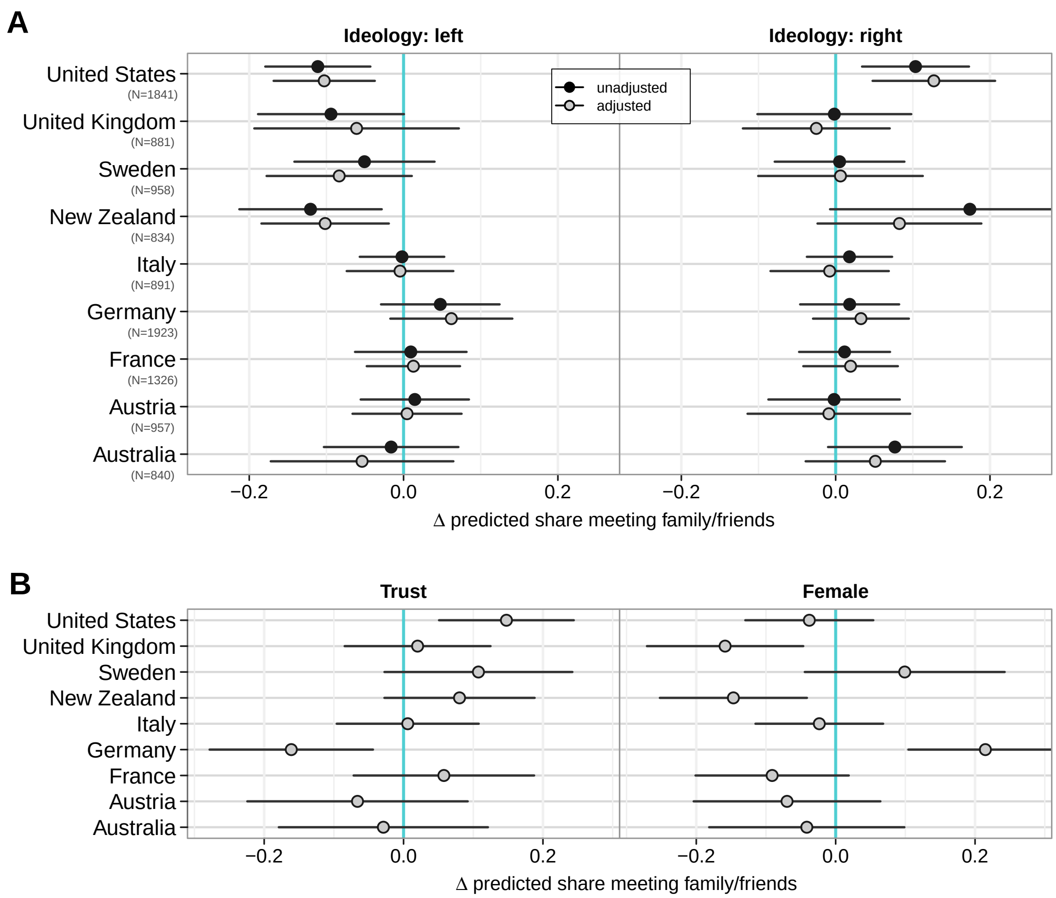

Figure I

Individual characteristics and probability of not following social distancing guidelines.

This figure plots changes in the predicted probability of non-compliance with 90% confidence intervals based on maximum likelihood estimates. Panel A plots the effect of a negative and positive standard deviation change in ideology. Panel B plots the effect of a change in indicator variables for trust and gender. Unadjusted models show the bivariate relationship between the two variables; adjusted models include a set of individual controls.

Panel A of Figure I displays the estimated relationship between political ideology and the probability of meeting friends and relatives despite social distancing guidelines for each country. It shows changes in the predicted probability of non-compliance (with 90% confidence intervals) arising from a standard deviation (SD) change in ideology away from the respective country mean. The first plot shows a SD decrease (i.e., moving a respondent to the left), while the second plot shows a SD increase (moving to the right). Our estimates suggest considerable cross-national variation in the impact of political ideology on following social distancing guidelines.

In the U.S., we find that individuals further on the right are more likely to meet friends or relatives during the pandemic. Substantively, our estimates suggests that individuals one SD to the right of the national mean are approximately 12% more likely to skirt social distancing. In contrast, individuals one SD to the left are approximately 10% more likely to follow social distancing guidelines.[6] This finding is consistent with evidence on polarized elite rhetoric over COVID-19 (Green et al. 2020) and evidence on partisan gaps in social distancing based on mobility patterns from smartphone data, spending behavior, and direct survey question (Allcott et al. 2020; Grossman et al. 2020; Kushner Gadarian et al. 2020; Painter and Qiu 2020). A qualitatively similar pattern exists in New Zealand and Australia, though it is less pronounced in the latter (and confidence intervals of the ideological gap always overlap zero).

The same pattern generally does not show up in the European countries. There, the estimated impact of ideology is approximately null. Following health guidelines is not a question of political ideology in the majority of countries under study. The absence of this association is not simply a mechanical result of very high or low compliance (indicating ceiling or floor effects). Countries like Austria and Germany show more variation in individual-level responses than the U.S. (Table III). The null result also emerges in different pandemic contexts, including large variation in COVID-19 deaths and strictness of lockdown-style policies. Our results are also not easily attributed to differences in the degree of party polarization across countries: levels of political polarization in the U.S. are quite similar to those in the other countries under study, whether one looks at mass polarization over the role of the government in the economy (Lindquist and Östling 2010) or the polarization of party positions (Lupu 2015: 343). What may be different in the U.S. is not polarization in general but the polarization of elite rhetoric on COVID-19 in particular (Green et al. 2020).

One might ask if the cross-country heterogeneity evident in our estimates is simply due to country idiosyncracies in political culture and if using more "basic" individual characteristics, such as gender, would produce estimates that are more consistent among countries. Panel B of Figure I thus plots changes in probabilities of non-compliance as function of gender and social trust. Previous work on COVID-19 finds gender to be a key predictor of (self-reported) willingness to follow social distancing (Barari et al. 2020; Galasso et al. 2020; Perrotta et al. 2020; Brouard et al. 2020). We further include interpersonal trust, because the literature on externalities and social dilemmas (Ostrom 2000) suggests its relevance for compliant behavior. An individual's compliance with non-pharmaceutical interventions contributes to the public good of containing a pandemic, but success depends on a large fraction of people changing their behavior. Under such a coordination dilemma, individuals with higher levels of interpersonal trust are more likely to behave cooperatively.

The estimates in panel B do not show higher levels of consistency for these additional variables across countries. The relationship between trust and non-compliance is positive in the U.S. and Sweden, but clearly negative in Germany. While women seem somewhat more likely to follow social distancing than men (the estimated adjusted gender gap is negative for 7 out of 9 countries), this gender difference is not statistically significant in most countries. Moreover, in Germany women are substantively less likely to comply with distancing guidelines.

4.2 Non-parametric Estimates of Ideology and Non-compliance

We now turn to a more flexible model linking political ideology to health behavior. The dependent variable in this analysis is the individual-level within-country estimate of the latent factor, θ_W, capturing the propensity to ignore health guidelines (thus, larger values represent a lower propensity to follow health guidelines). Denote by $\theta_{W,ir}$ the estimated latent variable value for individual $i(i = 1, \ldots, N_r)$ in region $r(r = 1, \ldots, R)$. For each of the nine countries in our sample we estimate the following model:

$$\theta_{W,ir} = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + f(z_i) + \xi_r + \epsilon_{ir},$$

where z_i represents political ideology, $\boldsymbol{x}_i$ is a vector of individual-level controls, including basic demographic characteristics, such as gender and age, and ξ_r represent unobserved regional-level differences.[7] We model region-specific constants as random effects drawn from a common normal distribution with freely estimated variance, $\boldsymbol{\xi} \sim N(0, \sigma_\xi^2)$.

It might be overly simplistic to constrain the relationship between health behavior and political ideology to be globally linear (or quadratic) *a priori* (e.g., Beck and Jackman 1998: 598). Instead we allow it to take on a flexible non-linear form if this is demanded by the data. We thus model f using B-splines (de Boor 1978; Hastie et al. 2017: 186). More precisely, over a sequence of L equidistant knot locations, we define L B-spline basis functions $B_l(z_i)$ with associated coefficients $(\gamma_1, \ldots, \gamma_L)$:[8]

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

Here, $B_l(\mathbf{z})$ is a set of basis functions and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)$ are basis coefficients representing their amplitudes. Identification of the estimated function is achieved by imposing some constraint of the form $\boldsymbol{A\gamma} = \mathbf{0}$. In our application, $\boldsymbol{A}$ imposes a centering constraint so that β_0 represents the "intercept" of the semi-parametric ideology regression. To control the "smoothness" of the fitted function (and to avoid overfitting), we impose a difference penalty term that punished too abrupt function jumps (Eilers and Marx 1996). In a Bayesian framework, this can be achieved by including a difference penalty in the prior distribution of the coefficients (cf. Brezger and Lang 2006)

$$\pi(\boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2\omega^2}\boldsymbol{\gamma}'\boldsymbol{K}\boldsymbol{\gamma}\right).$$

Here, $\boldsymbol{K}$ is a penalty matrix constructed via $\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{D}$, where $\boldsymbol{D}$ is a second-order difference matrix (created by twice applying the difference operator to adjacent spline coefficients; see appendix D.1 for an example), and ω^{-2} is a smoothness penalty term. The larger the penalty, the more our estimated function will be "shrunk" towards a linear fit (Hastie et al. 2017: 151). We provide more details in Appendix D.1 where we discuss the use of a scale-dependent hyperprior for ω adjudicating the trade-off between linearity and non-linearity in the estimated relationship, as well as our choice of priors for the remaining model parameters.

Table IV provides a comparative model assessment for a range of specifications. It provides the Watanabe-Akaike information criterion (Watanabe 2013), which penalizes model deviance by the effective number of parameters.[9] Specification A serves as reference point by modeling health behavior using only basic covariates: age, gender, education, and an indicator for being employed before the onset of the pandemic. Specification B adds ideology estimated using the semi-parametric approach described here (as well as the basic covariates included in A). A comparison of WAIC values shows that in every country the model including ideology is strongly preferred over the model with covariates only. The improvement in WAIC is about 300 on average and is greater than 100 in all countries but Austria (95) signifying that accounting for respondents' ideology improves model performance dramatically. The remaining specifications explore heterogeneity in the functional relationship between ideology and health behavior. We will return to them after discussing our core model results.

Table IV
Model fit comparisons. WAIC with effective number of parameters in italics.

Specification	AU	AT	FR	DE	IT	NZ	SE	UK	US
A: Basic demographics	2551	2435	4859	4964	2463	2428	2431	2408	4963
	7.2	*8.4*	*11.4*	*9.3*	*10.1*	*8.2*	*9.9*	*9.4*	*11.4*
B: f(Ideology)	2133	2340	4369	4766	2176	2063	2298	2089	4643
	10.3	*11.9*	*12.5*	*11.9*	*14.0*	*11.1*	*12.8*	*12.3*	*16.7*
C: Region differences	2134	2340	4370	4760	2176	2062	2299	2090	4644
	11.2	*15.1*	*18.0*	*26.1*	*15.7*	*13.9*	*14.4*	*13.6*	*20.2*
D: Government vote	2124	2339	3205	4751	2177	2061	2300	2090	4639
	11.7	*12.9*	*13.6*	*12.9*	*14.7*	*12.0*	*13.9*	*13.3*	*17.7*
E: Trust in executive	2127	2315	4367	4743	2170	2048	2300	2086	4638
	11.4	*13.1*	*13.4*	*13.0*	*15.2*	*12.2*	*13.8*	*13.5*	*17.3*

Note: $WAIC$ is the Widely Applicable Information Criterion of Watanabe (2010). WAIC penalizes model deviance by an estimate of the effective number of parameters (given in italics). Based on 10,000 MCMC samples. See online appendix D.3 for details.

Figure II plots semi-parametric estimates of the relationship between ideology and the health behavior adjustment latent variable. It plots expected values of the propensity to not follow health guidelines as a function of ideology, after partialling out basic individual characteristics and systematic regional differences. The variance of the dependent variable is fixed to unity in its generating IRT model and its mean is zero (cf. appendix C.1). Similarly, we scale ideology to mean zero and unit standard deviation in each country, so that both axes of Figure II can be interpreted in standard deviation units.

Figure II revels a substantial degree of effect heterogeneity among countries. There is a set of countries in which ideology has little to no relationship with not following health guidelines: France, Austria, and Sweden are prime examples. After accounting for the uncertainty in the estimated functional relationships (shown via 1,000 hairline function plots) changes in ideology do not significantly change the expected value of the outcome variable. This holds by and large for the United Kingdom as well, save for the fact that very left-leaning individuals (those more than one standard deviation below the national mean) are more likely to adjust their health behavior following expert guidelines. The estimated relationship for Germany reveals a modest (almost linear) trend: more conservative Germans are more likely to ignore health guidelines. It is however of limited magnitude (less than 1/10th of a standard deviation over the whole range of ideology).

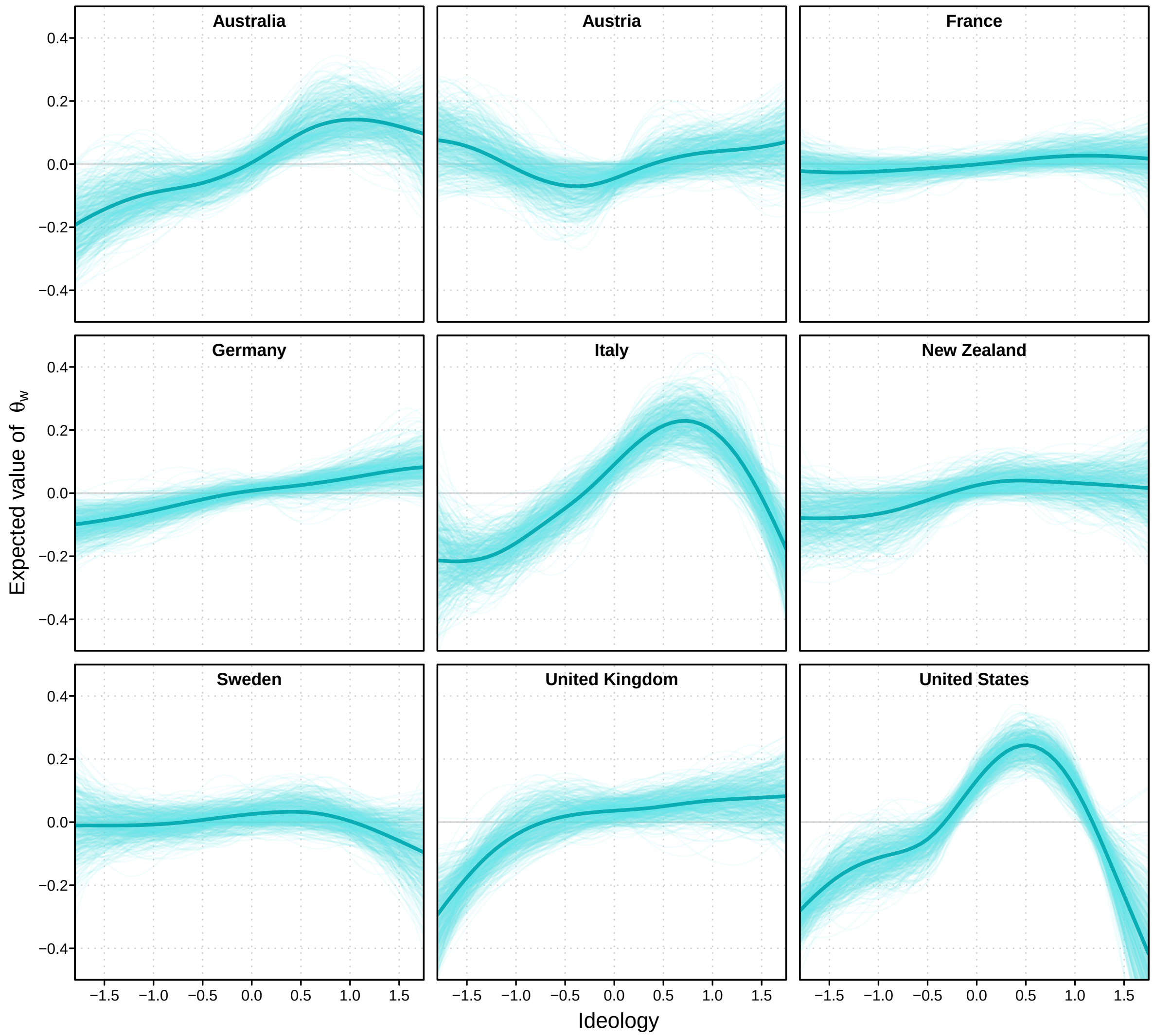

Figure II

Relationship between ideology and propensity to not follow health guidelines

This figure plots expected values of θ_W from semi-parametric ideology estimates based on a Bayesian mixed model representation of quadratic P-splines with second-order difference penalty while adjusting for individual characteristics and regional heterogeneity. The solid lines represents the posterior mean, estimation uncertainty is represents by hairlines plotting 500 function evaluations calculated using a random subsample of the MCMC output.

In another group of countries, most notably Australia, the U.S., and Italy, ideology is strongly related to the propensity to not change ones health behavior. In the U.S., for an individual placing herself at the average of the political spectrum, her propensity to not change her behavior is already significantly different from 0 (the overall mean of the latent variable). Moving an individual living in the U.S. half a standard deviation to the right of the national mean ideology, increases their propensity to ignore health guidelines by about 0.2 standard deviations. A commensurate change in Italy shows a similar effect. The magnitude of the relationship is somewhat more muted in Australia, but it is still significant both in the statistical and substantive sense.[10]

Geography and ideology It is germane to ask if our analyses are affected by systematic sub-national geographic differences, such as regional or state-level differences in civic traditions (Putnam 2000), or variation in state-level ideology (Rabinowitz et al. 1984). While our preferred model does include region or state random effects, a more extended specification allows the shape of the estimated ideology function to vary across geographical units. However, this flexibility comes at the cost of a large increase in the number of model parameters. The fit statistics for specification C in Table IV take this into account (the number of effective parameters is given in italics). Remarkably, this increased flexibility does not, in general, lead to a notable improvement in WAIC. Germany is an instructive case. While the number of effective parameters in specification C almost doubles, the differences in WAIC over specification B is only 6. In the remaining countries the fit criterion remains effectively unchanged or even worsens. We conclude from this analysis that ideology is a major source of heterogeneity among individuals and that after accounting for the possibly nonlinear function form of ideology additional regional effect heterogeneity is of no or limited importance.

Support for the government A possibility, neglected so far in our analysis, is that ideological patterns we find are driven in part by respondents' alignment with the current government. Respondents who chose the current government in the last election, or who currently still favor it, might be more receptive to its messaging and more likely to follow health guidelines. Conversely, in settings where a government downplays the risk of the pandemic, receptive respondents might be more likely to not follow expert health guidelines. The latter relationship might explain the ideological pattern found in the U.S., where the political polarization of pandemic misinformation was high and partly linked to executive communication (Havey 2020; Kushner Gadarian et al. 2020).

In specifications D and E in Table IV we show results from two models where we adjust for the fact that the expected value of changing health behavior might depend on respondents' support of the government. We estimate two sets of models. The first (specification D) includes an indicator variable equal to one if a respondent cast a vote for the governing party (or coalition) in the last election. The second (E) includes a variable capturing respondents' trust in the head of the executive.[11] It is measured on a four-point scale with response options ranging from "don't trust at all' to "trust completely". While both specification lead to a clear improvement in WAIC in a few countries (France and Germany in specification D; Austria and Germany in E), most seem rather unaffected by this change. The left panel of Figure III plots estimates for the coefficient on the government vote indicator variable for each country (based on estimates reported in Table D.1). It reveals that support for the government does not significantly influence the propensity to ignore health guidelines in most countries, with the exception of the U.S., Australia, and Germany. For the latter two, supporting the government goes hand in hand with a lower propensity to resist health advice. The U.S. is the only country in our nine-nation sample in which support for the government is associated with a significantly *higher* propensity of not following health advice. The right panel of Figure III plots estimates for trust in the head of the executive (estimates represent a standard deviation increase). In general, higher trust is related to a lowered propensity of ignoring health guidelines (with small to zero coefficients in France, Italy, and Sweden). Again, the U.S. emerges as the exception showing the opposite relationship.

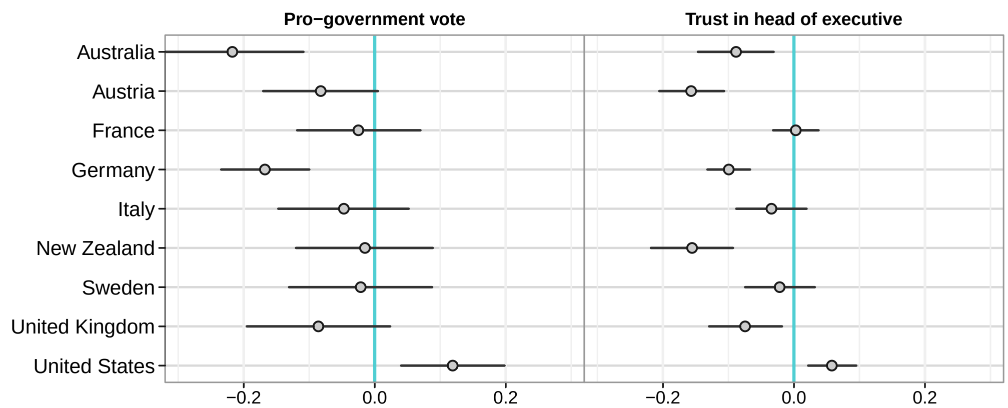

Figure III
Estimates of respondents' government support on health behavior

This figure plots changes in expected values of θ_W for respondents who voted for the currently governing party or coalition in the last election (left panel) and for a standard deviation increase of respondents' trust in the head of the executive (right panel). Posterior means and 90% credible intervals based on 10,000 MCMC samples.

While support of, and trust in, the executive is clearly important in some contexts, it does not alter significantly the functional form of the impact of ideology on health behavior. Figure D.2 in the appendix compares estimated functions of ideology after accounting for government support with those displayed in Figure II above, and reveals no differences of substantive importance. Finally, in appendix D.3 we further explore the role of respondents' trust in experts (scientists and doctors). We find that, in general, higher trust does lower the propensity to not follow health advice. However, accounting for trust in experts has little impact on the estimated impact of ideology.

5 Relation to pandemic severity and policy

We end our empirical analysis with an exploratory investigation of the relationship between our estimates of a populations' propensity to not follow health guidelines (estimated from the list experiment and the latent variable model). This exploratory analyses, summarized in Figure IV, provide two empirical insights. We find that the estimated share of individuals meeting family and friends, as well as the country-level aggregate of the latent variable, are negatively correlated with the total COVID-19 related deaths in the week prior to the survey (per million inhabitants). The data we use are official government-reported counts compiled by researchers at Johns Hopkins University (Dong et al. 2020).[12] On average, countries with lower reported deaths, like Austria or the U.S., exhibit significantly higher levels of non-compliance than countries with higher reported deaths, such as France and Italy.

The right two panels of Figure IV show a similar negative association with the stringency of the government response to the pandemic. This finding is consistent with country-level results from the U.S. (Painter and Qiu 2020). The stringency index is taken from the Oxford COVID-19 Government Response Tracker (Hale et al. 2020) and measures (on a scale from 0 to 100) the strictness of lockdown-style policies, such as restrictions of movement and school closures, that

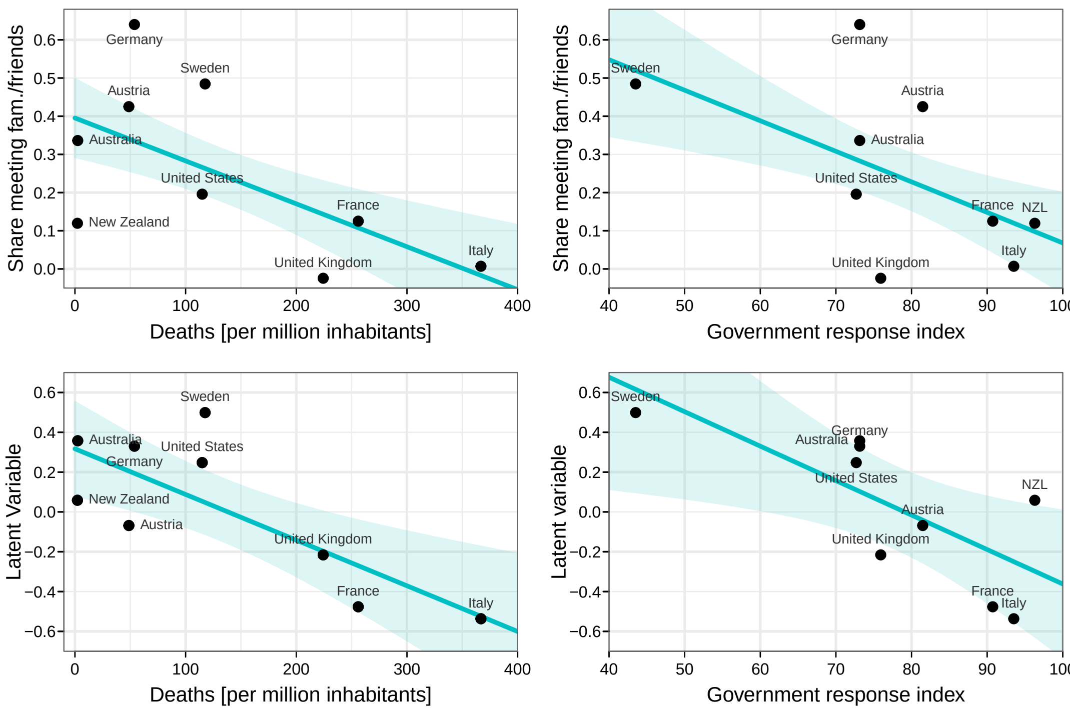

Figure IV
Relationship between Covid-19 related deaths, lockdown strictness and health behavior estimates
The upper two panels of this figure plot the fraction of individuals meeting family and friends estimated from the list experiment (y-axis) against total COVID-19 related deaths in the week before the survey and the strictness of lockdown-style measures. The lower two panels show the same relationship but use on the y-axis the within and between component of our latent health measure ($\theta_w + \theta_b$). Each plot also includes a linear fit from a robust regression (using an M-estimator with Huber objective function) and 90% confidence intervals.

primarily restrict individuals' behavior. On average, countries with strict policy measures, such as France or Italy, show lower levels of ignorance of health guidelines compared to countries with more lenient measures, such as Australia, the U.S., or (representing an extreme case) Sweden. We hasten to add that this analysis is merely descriptive. Nonetheless, it helps to make sense of the variation uncovered by the measurement experiment and latent variable and is broadly consistent external monitoring and sanctions being relevant factors (Luttmer and Singhal 2014). Fruitful future work could employ within-country variation in pandemic severity on policy (for example, in U.S. states) to put these relationships to a stricter test.

6 Discussion and Conclusion

Compared to direct survey questions, the list experimental approach makes our findings less susceptible to measurement errors induced by social desirability bias that have been raised in the literature. A potential limitation of our experiment is its external validity—understood as the scope of actions captured by the experiment, which focuses on meeting friends or relatives. To address this limitation, we have also constructed and analyzed a latent variable capturing health behavior adjustment on a broader scope. Our conclusions about the country-varying relationship between ideology and compliance are supported by both approaches.

While there is a clear ideological gap in compliance with health guidelines during the COVID-19 pandemic in the U.S., in a majority of the countries we studied there is no comparable association between ideology and compliance. Our results highlight the importance of taking a comparative perspective. The degree of heterogeneity revealed in both the list experiment and the semiparametric latent variable analyses, suggests that researchers studying the pandemic should be cognizant of the pitfalls when extrapolating from both single-country studies and standard pooled (or homogenous) country-regressions.[13]

An implication of our results is that it may be difficult for policy-makers to learn from other countries' experiences when crafting policies intended to enhance compliance with public health guidelines. While behavioral social science can draw on a repertoire of experimentally tested 'nudges' to enhance compliance (Bavel et al. 2020), our results highlight that the social and political characteristics of individuals less likely to follow health guidelines vary across countries. Thus, behavioral interventions intended to target non-compliers should not be based on the assumption that "non-compliers" behave identically across countries. Given heterogeneity in ideology, otherwise observationally identical individuals might make very different choices when confronted with expert health messaging. Clearly, more context-specific evidence is needed.

A somewhat more encouraging aspect of our findings is that while the relevance of political ideology for social distancing is pronounced in some countries—including the U.S.—this is by no means the rule. At least during the initial lockdown stage of the pandemic, the same was not true in Europe with the exception of Italy, possibly giving policymakers more scope for (future) action.

Notes

[1]For example, research on self-reported behavior and attitudes shows voter turnout and racial animus estimates to be affected by social desirability bias (e.g., Belli et al. 2001; Bernstein et al. 2001; Kuklinski et al. 1997).

[2]Blair and Imai (2012) list applications in a wide range of fields. Studying compliance during the COVID-19 pandemic, single-country list experiments have been conducted (Larsen et al. 2020; Munzert and Selb 2020).

[3]Our hypothesis regarding ideology as well as the analysis of other factors discussed later was pre-registered. See `http://aspredicted.org/blind.php?x=hv7yv2`.

[4]Note that we do not claim that the structure of political ideology in Western mass publics is one-dimensional. A large body of work provides evidence of (at least) a second dimension (see, among many, Heath et al. 1994; Treier and Hillygus 2009). We use our simple measure as a first-order approximation to individual differences in political beliefs and attitudes beyond simple demographic categories.

[5]Trust is measured by an indicator variable equal to 1 for respondents agreeing that "most people can be trusted" (0 if "you can never be too careful when dealing with other people"). Subjective health is measured by the item: "Generally speaking, would you say that your health is [Very good/Good/Quite good/Bad/Very bad]?"

[6]This finding is only partly consistent with our pre-registered hypothesis on the role of ideology. We expected that individuals with more extreme political preferences (on the left and right) would be less likely to comply (cf. Brouard et al. 2020) implying a U-shaped pattern. Our semi-parametric analysis below provides more evidence that counters this initial hypothesis.

[7]These represent NUTS regions in European countries and states in Australia and the U.S.

[8]We use quadratic B-spline bases with 10 knots penalized via second-order difference penalties and collapse the outer two categories of the ideology scale for increased numerical stability. B-splines are constructed by augmenting the 10 equally-spaced interior knots (spanning all values of ideology) with 2 upper and 2 lower boundary knots.

[9]WAIC can be seen as an extension of the AIC. It is more useful in our context since it accounts for the *effective* number of parameters to adjust for overfitting. The number of effective parameters will usually be much lower than the number of coefficients in the model due to the penalization employed in the estimation. See appendix D.4 for more discussion of its merits and an alternative fit measure.

[10]Italy and the U.S. also show a decrease in non-compliance at the very upper end of the left-right ideological spectrum. This is not an artefact of the spline construction (such as the choice of spline basis or knot placements). The mean of the latent factor for respondents placing themselves in the highest category is indeed lower than the next lowest category. One might hypothesize that individuals in the rightmost individual category might be predominantly older and more at risk. But note, that our estimates are already adjusted for age, gender, and education differences. A more systematic exploration of this pattern (including its cross-national component) merits further research.

[11]We use the latter to capture the fact that respondents' support during the pandemic might differ from what one would infer from their vote in the past. For an analysis of leader support during the pandemic see Freden and Sikstrom in this issue.

[12]While reporting standards vary across countries, these data have been widely reported in the media and thus shaped the public salience of the pandemic and its associated risks.

[13]We thank an anonymous reviewer for pointing this out to us. Figure D.5 in the appendix illustrates this point graphically. There we compare our results to what one would obtain when using a common specification, namely a (pooled) linear model with country fixed effects. While such a model yields a significant slope for ideology ($\beta = 0.046 \pm 0.011$) and describes the ideology-health relationship very well in some countries (e.g., in Germany), it provides a misleading sense for many others (such as France, the U.S., or Italy).

References

Allcott, H., L. Boxell, J. C. Conway, M. Gentzkow, M. Thaler, and D. Y. Yang (2020, April). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. Technical Report 26946, National Bureau of Economic Research.

Andersen, M. (2020). Early evidence on social distancing in response to covid-19 in the united states. *SSRN, https://dx.doi.org/10.2139/ssrn.3569368*.

Barari, S., S. Caria, A. Davola, P. Falco, T. Fetzer, S. Fiorin, ..., and F. R. Slepoi (2020). Evaluating covid-19 public health messaging in italy: Self-reported compliance and growing mental health concerns. *medRxiv*.

Barrios, J. M. and Y. Hochberg (2020, April). Risk perception through the lens of politics in the time of the covid-19 pandemic. Working Paper 27008, National Bureau of Economic Research.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior 24*(2), 117–150.

Bavel, J. J. V., K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, ..., and R. Willer (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour 4*(5), 460–471.

Beck, N. and S. Jackman (1998). Beyond Linearity by Default: Generalized Additive Models. *American Journal of Political Science 42*(2), 596–627.

Belli, R. F., M. W. Traugott, and M. N. Beckmann (2001). What leads to voting overreports? contrasts of overreporters to validated voters and admittednonvoters in the american national election studies. *Journal of Official Statistics 17*(4), 479–498.

Bernstein, R., A. Chadha, and R. Montjoy (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly 65*(1), 22–44.

Blair, G., W. Chou, K. Imai, B. Park, and A. Coppock (2020). *list: Statistical Methods for the Item Count Technique and List Experiment*. R package version 9.2.

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

Brezger, A. and S. Lang (2006). Generalized structured additive regression based on Bayesian P-Splines. *Computational Statistics and Data Analysis 50*, 967–991.

Brouard, S., P. Vasilopoulos, and M. Becher (2020). Sociodemographic and psychological correlates of compliance with the covid-19 public health measures in france. *Canadian Journal of Political Science 52*(2), 253–258.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes (1960). *The American Voter*. Chicago: John Wiley & Sons.

Converse, P. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent*, pp. 206–261. New York: Free Press.

Daoust, J.-F., R. Nadeau, R. Dassonneville, E. Lachapelle, E. Bélanger, J. Savoie, and C. van der Linden (2020). How to survey citizens' compliance with COVID-19 public health measures? evidence from three survey experiments. *Journal of Experimental Political Science forthcoming*.

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

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.

Droitcour, J., R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati (2011). The item count technique as a method of indirect questioning: A review of its development and a case study application. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement errors in surveys*, pp. 185–210. John Wiley & Sons.

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

Feldman, S. (2013). Political ideology. In L. Huddy, D. O. Sears, and J. S. Levy (Eds.), *The Oxford Handbook of Political Psychology*, pp. 591–626. Oxford: Oxford University Press.

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: Routeledge.

Galasso, V., , V. Pons, P. Profeta, M. Becher, S. Brouard, and M. Foucault (2020). Gender differences in covid-19 attitudes and behavior: Panel evidence from eight countries. *Proceedings of the National Academy of Sciences 117*(44), 27285–27291.

Gerber, A. S. and G. A. Huber (2009). Partisanship and economic behavior: Do partisan differences in economic forecasts predict real economic behavior? *The American Political Science Review 103*(3), 407–426.

Graham, M. H. and M. W. Svolik (2020). Democracy in america? partisanship, polarization, and the robustness of support for democracy in the united states. *American Political Science Review 114*(2), 392–409.

Green, J., J. Edgerton, D. Naftel, K. Shoub, and S. J. Cranmer (2020). Elusive consensus: Polarization in elite communication on the covid-19 pandemic. *Science Advances*.

Grossman, G., S. Kim, J. M. Rexer, and H. Thirumurthy (2020). Political partisanship influences behavioral responses to governors' recommendations for covid-19 prevention in the united states. *Proceedings of the National Academy of Sciences 117*(39), 24144–24153.

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.

Hastie, T., R. Tibshirani, and J. Friedman (2017). *The Elements of Statistical Learning. 2nd edition.* New York.

Havey, N. F. (2020). Partisan public health: how does political ideology influence support for covid-19 related misinformation? *Journal of Computational Social Science 3*(2), 319–342.

Heath, A., G. Evans, and J. Martin (1994). The measurement of core beliefs and values: The development of balanced socialist/laissez faire and libertarian/authoritarian scales. *British Journal of Political Science 24*(1), 115–132.

Holbrook, A. L. and J. A. Krosnick (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly 74*(1), 37–67.

Huber, J. and R. Inglehart (1995). Expert interpretations of party space and party locations in 42 societies. *Party Politics 1*(1), 73–111.

Huijts, T., J. M. Perkins, and S. V. Subramanian (2010). Political regimes, political ideology, and self-rated health in europe: a multilevel analysis. *PloS One 5*(7), e11711.

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association 106*(494), 407–416.

Jost, J. T., J. Glaser, A. W. Kruglanski, and F. J. Sulloway (2003). Political conservatism as motivated social cognition. *Psychological Bulletin 129*(3), 339–375.

Kuklinski, J. H., M. D. Cobb, and M. Gilens (1997). Racial attitudes and the "new south". *The Journal of Politics 59*(2), 323–349.

Kushner Gadarian, S., S. W. Goodman, and T. B. Pepinsky (2020, March). Partisanship, health behavior, and policy attitudes in the early stages of the covid-19 pandemic. Working paper, SSRN.

Larsen, M. V., M. B. Petersen, and J. Nyrup (2020, Apr). Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias? Psyarxiv.

Levi, M. and L. Stoker (2000). Political trust and trustworthiness. *Annual Review of Political Science 3*, 475–507.

Lindquist, E. and R. Östling (2010). Political polarization and the size of government. *American Political Science Review 104*(3), 543–565.

Lupu, N. (2015). Party polarization and mass partisanship: A comparative perspective. *Political Behavior 37*(2), 331–356.

Luttmer, E. F. and M. Singhal (2014). Tax morale. *The Journal of Economic Perspectives 28*(4), 149–168.

Miller, J. D. (1984). A new survey technique for studying deviant behavior. PhD thesis. George Washington University.

Munzert, S. and P. Selb (2020). Can we directly survey adherence to non-pharmaceutical interventions? evidence from a list experiment conducted in germany during the early corona pandemic. *Survey Research Methods 14*(2), 205–209.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives 14*(3), 137–158.

Painter, M. and T. Qiu (2020, March). Political beliefs affect compliance with COVID-19 social distancing orders. Technical report, SSRN.

Perrotta, D., A. Grow, F. Rampazzo, J. Cimentada, E. Del Fava, S. Gil-Clavel, and E. Zagheni (2020). Behaviors and attitudes in response to the covid-19 pandemic: Insights from a cross-national facebook survey. *medRxiv*.

Putnam, R. (2000). *Bowling Alone. The collapse and revival of american community*. New York: Simon and Schuster.

Rabinowitz, G., P.-H. Gurian, and S. E. Macdonald (1984). The structure of presidential elections and the process of realignment, 1944 to 1980. *American Journal of Political Science 27*, 611–35.

Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society B 41*(1), 40–45.

Schaller, M. (2011). The behavioural immune system and the psychology of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences 366*(1583), 3418–3426.

Schaller, M. and J. H. Park (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science 20*(2), 99–103.

Short, S. E. and S. Mollborn (2015). Social determinants and health behaviors: conceptual frames and empirical advances. *Current Opinion in Psychology 5*, 78–84.

Subramanian, S. and J. M. Perkins (2009, 03). Are republicans healthier than democrats? *International Journal of Epidemiology 39*(3), 930–931.

Subramanian, S. V., T. Huijts, and J. M. Perkins (2009, 06). Association between political ideology and health in Europe. *European Journal of Public Health 19*(5), 455–457.

Treier, S. and D. S. Hillygus (2009). The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly 73*(4), 679–703.

Utych, S. M. and L. Fowler (2020). Age-based messaging strategies for communication about COVID-19. *Journal of Behavioral Public Administration 3*(1).

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.

West, R., S. Michie, G. J. Rubin, and R. Amlôt (2020). Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nature Human Behaviour 4*(5), 451–459.

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

Contents

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.[1] 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).[2]

---

[1] The 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).

[2] In 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.

Table A.1
Survey details

	Fieldwork	Sample size	Resp. rate[a]	Completion rate[a]	Deaths[b]	Gov. response index[c]
Australia	04/15 - 04/19	1 007	0.10	0.76	3	73.2
Austria	04/15 - 04/18	1 000	0.33	0.95	49	81.5
France	04/15 - 04/16	2 020	0.47	0.96	256	90.7
Germany	04/16 - 04/18	2 000	0.31	0.93	54	73.2
Italy	04/15 - 04/17	997	0.37	0.94	367	93.5
New Zealand	04/15 - 04/18	998	0.38	0.94	2	96.3
Sweden	04/16 - 04/18	1 009	0.33	0.95	118	38.0
United Kingdom	04/15 - 04/17	1 000	0.35	0.94	224	75.9
United States	04/15 - 04/20	2 007	0.12	0.68	114	74.5

a Response rate S/I, completion rate $C/(S-Q)$; I is the number of individuals invited, S the number of started surveys, Q number of surveys removed due to quota being fulfilled, C number of completed surveys.
b Median number of deaths per million inhabitants in week prior to survey. Source: COVID-19 Data Repository, Center for Systems Science and Engineering, JHU (Dong et al. 2020).
c Government response stringency index. Source: Oxford Covid-19 Government Response Tracker (Hale et al. 2020).

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 seizable 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

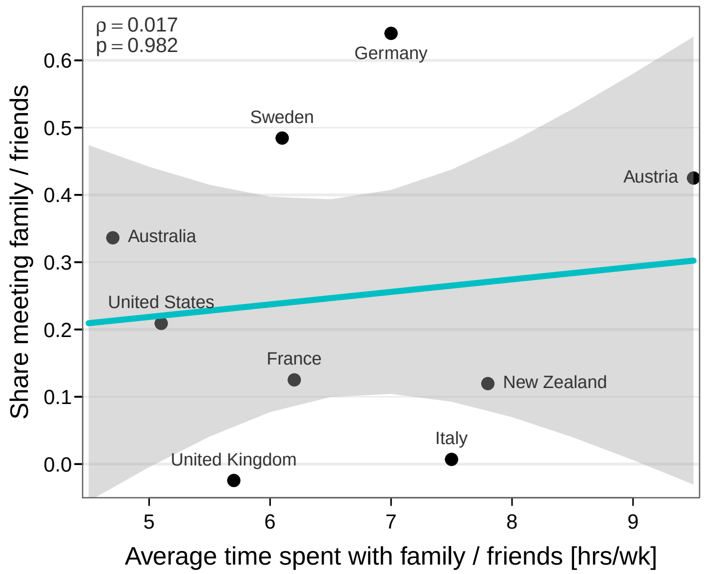

Figure B.1
Social ties and prevalence of noncompliance

This figure shows that existing patterns of social interactions are not strongly related to the share of individuals not following health guidelines during the pandemic. It plots the average time spent socializing with family or friends [as primary activity, in hours/week] around 2018 (data from OECD 2020) against experimental estimates of the share of individuals meetings family or friends despite health guidelines in 2020. Robust regression line with confidence bands superimposed. The Spearman rank correlation between both measures is 0.017 with a p-value of 0.982.

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.[3] 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

[3]Note 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.

					Prop. zeros		Design effect	
Country	N_0	N_1	Y_0 avg.	Y_0 CV	Y_0	Y_1	p	p_{BH}
Australia	513	494	1.395	0.823	0.23	0.22	1.000	1.000
Austria	498	498	1.061	1.014	0.37	0.27	0.915	1.000
France	1014	1006	0.696	1.187	0.48	0.45	0.984	1.000
Germany	1000	1000	1.081	1.009	0.36	0.12	0.098	0.443
Italy	498	499	0.753	1.318	0.53	0.54	0.616	1.000
New Zealand	500	498	1.131	0.681	0.16	0.15	1.000	1.000
Sweden	502	507	0.912	1.148	0.42	0.29	1.000	1.000
United Kingdom	500	500	1.244	0.757	0.18	0.24	0.021	0.192
United States	949	1006	1.125	1.001	0.36	0.33	1.000	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 p-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

	Control			Treated			
	Mean (1)	s.e. (2)	s.d. (3)	Mean (4)	s.e. (5)	s.d. (6)	Diff (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	−0.218
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	0.53	0.02	0.50	0.033
Ideology	4.91	0.11	2.30	4.86	0.10	2.14	−0.048
Trust	0.50	0.02	0.50	0.51	0.02	0.50	0.012
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, τ, (referred to as "difficulties" in the IRT literature) and coefficients, λ, 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 the model takes the form $y_i = \tau + \lambda f_i + \epsilon_i$, where the distribution of residuals ϵ 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.

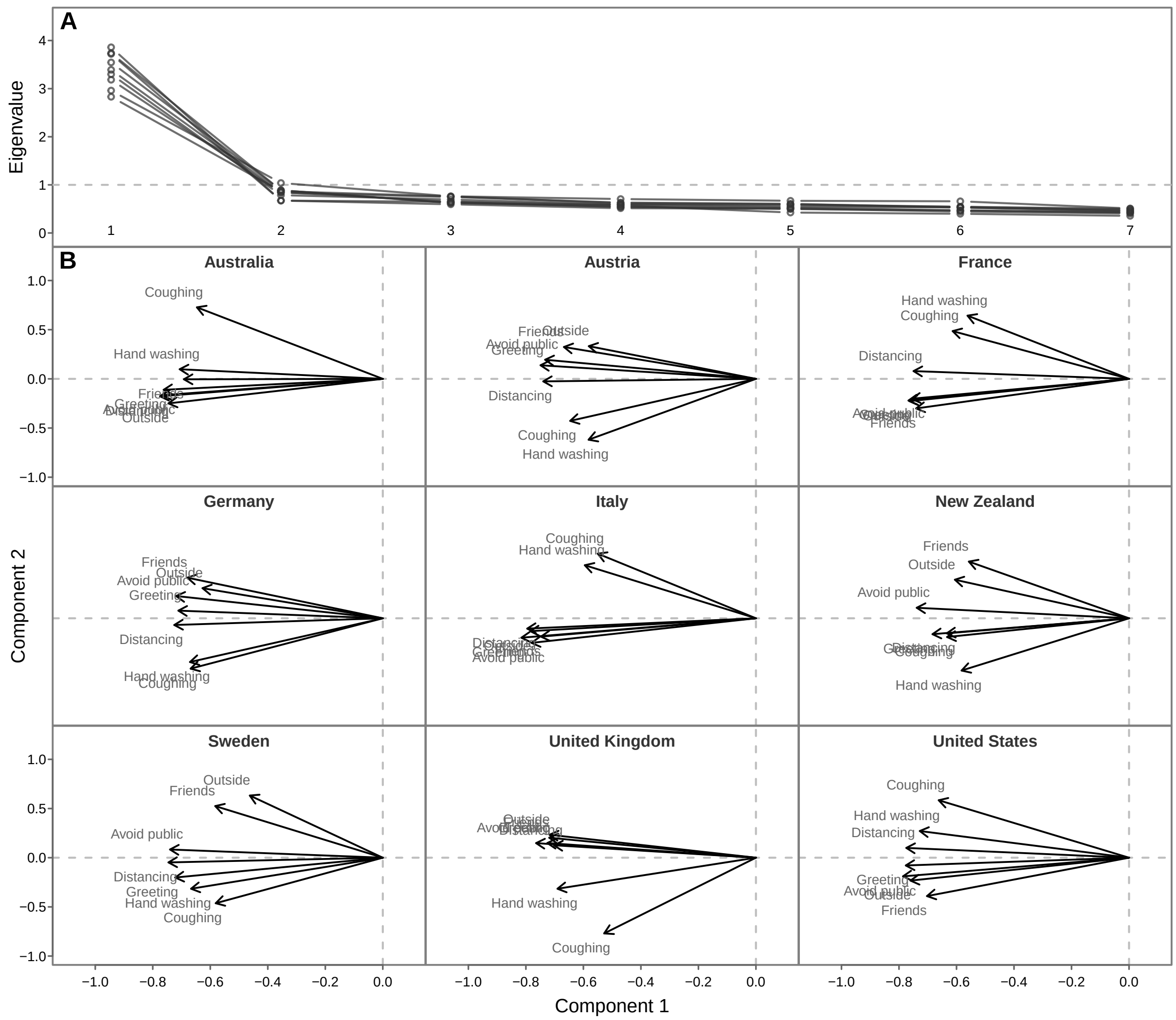

Figure C.1
Nonlinear principal components analysis of behavioral adjustment battery.

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^*_k as being driven by an underlying latent factor θ_W. We estimate the following measurement system

$$\begin{aligned} y_{ij1} &= \tau_{j1} + \lambda_{j1}\theta_{W_{ij}} + \epsilon_{ij1} \\ &\vdots \\ y_{ijK} &= \tau_{jK} + \lambda_{jK}\theta_{W_{ij}} + \epsilon_{ijK} \end{aligned} \tag{C.1}$$

where τ are item intercepts, λ are factor loadings and $\theta_{W_{ij}}$ is a latent factor representing individual propensity to change behavior. For identification, $\theta_{W_{ij}} \sim N(0, 1)$ as in standard IRT models.[5] Residuals ϵ_{ijk} are also called uniqenesses 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:

$$\tau_{jk} = \tau_k + \lambda\theta_{B_j} + \zeta_{\tau_{jk}} \tag{C.2}$$

$$\lambda_{jk} = \lambda_k + \lambda\theta_{\psi_j} + \zeta_{\lambda_{jk}} \tag{C.3}$$

$$\theta_B \sim N(0, \sigma_B^2) \tag{C.4}$$

$$\theta_{\psi_j} \sim N(0, \sigma_\psi^2) \tag{C.5}$$

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

Fox (2010) discusses the identification constraints needed to separately identify varying factor means and variances with both intercepts and leadings hierarchically modeled. We follow the strategy outlined in Asparouhov and Muthén (2015). Note that the loadings λ are equal for θ_{B_j} and θ_{ψ_j}.

The variation in item intercepts over countries is captured by σ_τ^2 while the variation in loadings is captured by σ_λ^2. The systematic country-variation of individual factor means is captured by σ_B^2; the variation in the factor variance is captured by σ_ψ^2.

The model is estimated using Gibbs sampling using latent data augmentation for the dichotomous variables. We specify normal priors for all λ and τ 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 $V = diag(2)$ and degrees of freedom set to $\nu = dim(V) + 1 = 3$.

C.2. Estimates and comparison to list experiment

Table C.1 shows estimates for all model parameters. Columns τ and λ represent means of the threshold and loading random coefficients. Displayed are estimates (posterior means) with 95% credible intervals in brackets. The σ_τ^2 and σ_λ^2 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, θ_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	τ_k	$\sigma^2_{\tau_k}$	λ_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]	
σ^2_B			0.153	
σ^2_ψ			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, τ, and loadings/discrimination parameters, λ. Columns σ^2 show corresponding variances of country-specific deviations. Prior on latent factor variances is $IW(\boldsymbol{I}_2, df = 3)$; priors on item parameters are $\pi(\boldsymbol{\tau}) \sim N_K(0, s\boldsymbol{I}_K)$ and $\pi(\boldsymbol{\tau}) \sim N_K(0, s\boldsymbol{I}_K)$ 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

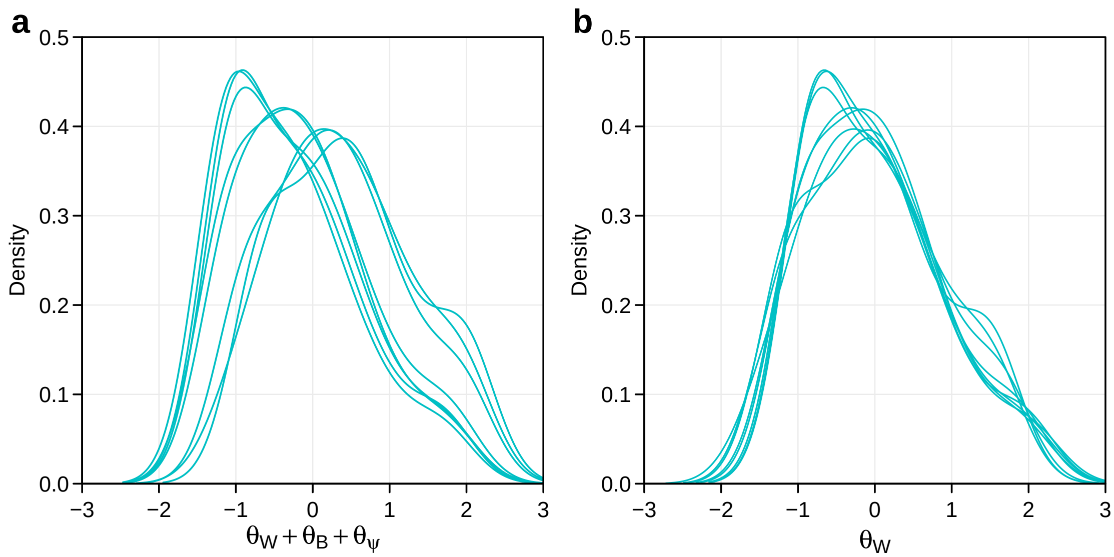

Figure C.2
Latent components of non-compliance with health guidelines

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 θ_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 Mean difference	Latent factor RC-IRT
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:

$$\theta_{W,ir} = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + f(z_i) + \xi_r + \epsilon_{ir}, \tag{D.1}$$

with the flexible function of ideology approximated via

$$f(\mathbf{z}) = \sum_{l=1}^{L} \gamma_l B_l(\mathbf{z}). \tag{D.2}$$

The B-spline coefficients $\boldsymbol{\gamma}$ are penalized using a quadratic penalty via a smoothing parameter λ:

$$\lambda^* \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma}, \tag{D.3}$$

where the positive (semi-)definite precision matrix $\boldsymbol{K} = \boldsymbol{D_r'}\boldsymbol{D_r}$ and $\boldsymbol{D_r}$ is a matrix of r-th order differences (Eilers and Marx 1996). It is created by applying a r-th order difference operator Δ^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.

$$\mathbf{D}_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \qquad \mathbf{K} = \mathbf{D}_1'\mathbf{D}_1 \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

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 $\boldsymbol{\gamma}$ is then given by:

$$p(\boldsymbol{\gamma}|\omega^2) \propto \left(\frac{1}{\omega^2}\right)^{\frac{rk(\mathbf{K})}{2}} \exp\left(-\frac{1}{2\omega^2}\boldsymbol{\gamma}'\mathbf{K}\boldsymbol{\gamma}\right) \tag{D.4}$$

and $\lambda^* = 1/\omega^2$.

The "smoothness" of the estimated function is thus influenced directly by ω^2. In a Bayesian framework, we learn about ω^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:

  $$d(p||p_0) = \sqrt{2 * KLD(p||p_0)} \tag{D.5}$$

  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 υ specified by the rate of decay r defined above (cf. Klein et al. 2016 Theorem 1 and appendix A.1)

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

We use this prior in our specifications used in the main text, with υ 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 ω (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 $\sigma^2_\xi \sim G^{-1}(a_0, b_0)$.

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

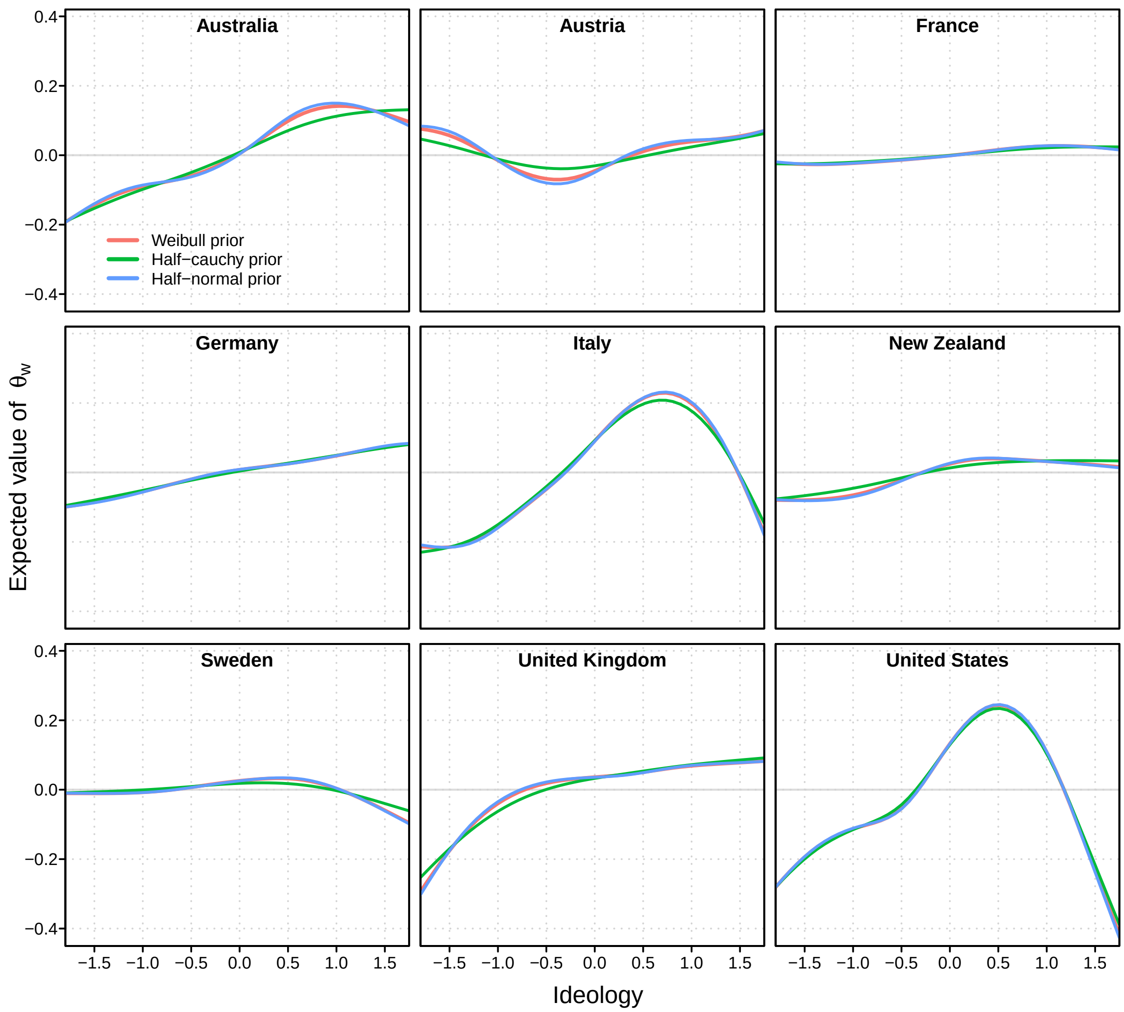

Figure D.1
Comparison of different smoothing variance prior choices

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 δ. 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

	AUS	AUT	FRA	DEU	ITA	NZL	SWE	UK	USA
Specification A									
$dim(\boldsymbol{\gamma})$	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
ω^2	0.0026	0.0031	0.0003	0.0007	0.0048	0.0012	0.0007	0.0020	0.0051
σ^2_ξ	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)
$dim(\boldsymbol{\gamma})$	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
ω^2	0.0035	0.0032	0.0006	0.0008	0.0046	0.0010	0.0007	0.0018	0.0051
σ^2_ξ	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	891	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)
$dim(\boldsymbol{\gamma})$	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
ω^2	0.0029	0.0037	0.0004	0.0008	0.0047	0.0010	0.0007	0.0020	0.0053
σ^2_ξ	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: δ 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. $dim(\boldsymbol{\gamma})$ indicates the number of spline coefficients, EDF is the estimated (posterior mean) effective degrees of freedom. ω is the posterior mean of the inverse smoothing parameter in the spline coefficient prior. σ^2_ξ 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.

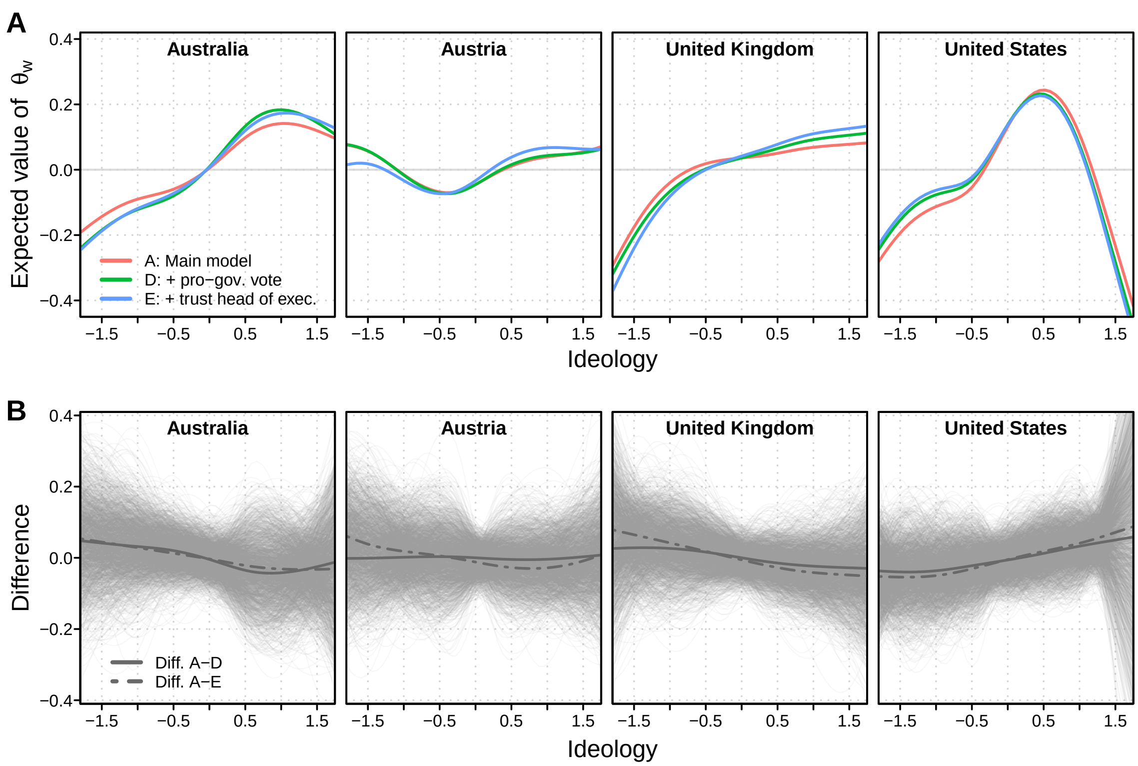

Figure D.2
Impact of ideology when adjusting for government support

Panel A plots expected values of θ_W from our main model (A) and specifications adjusting for past pro-government vote (D) and trsut 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 to 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.

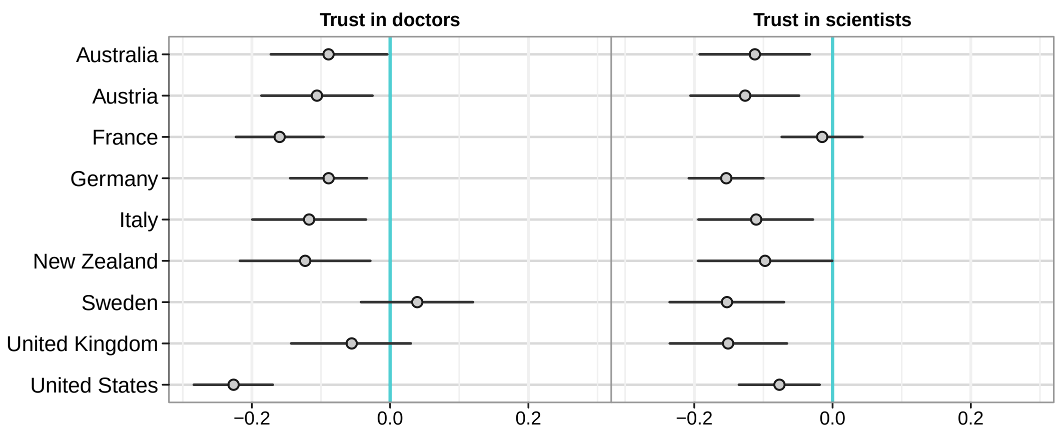

Figure D.3
Estimates of respondents' trust in experts on health behavior

This figure plots changes in expected values of θ_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.

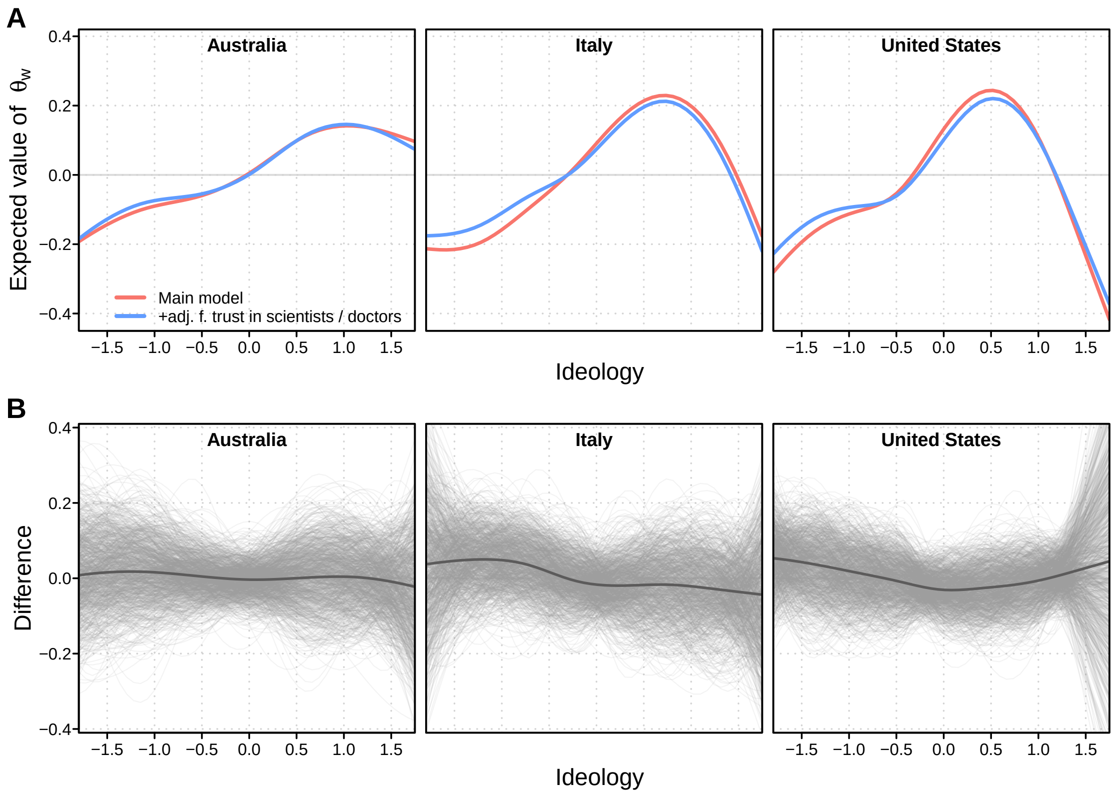

Figure D.4
Impact of ideology when adjusting for trust in doctors and scientists
Panel A plots expected values of θ_W from our main model (A) and a specification including for two variables capturing respondents' level of trust in doctors and scientists. Panel B plots differences with uncertainty represented by differences of 1,000 randomly drawn function evaluations.

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.

Table D.2
Model comparisons. DIC with effective number of parameters in italics.

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*	*13.3*	*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).

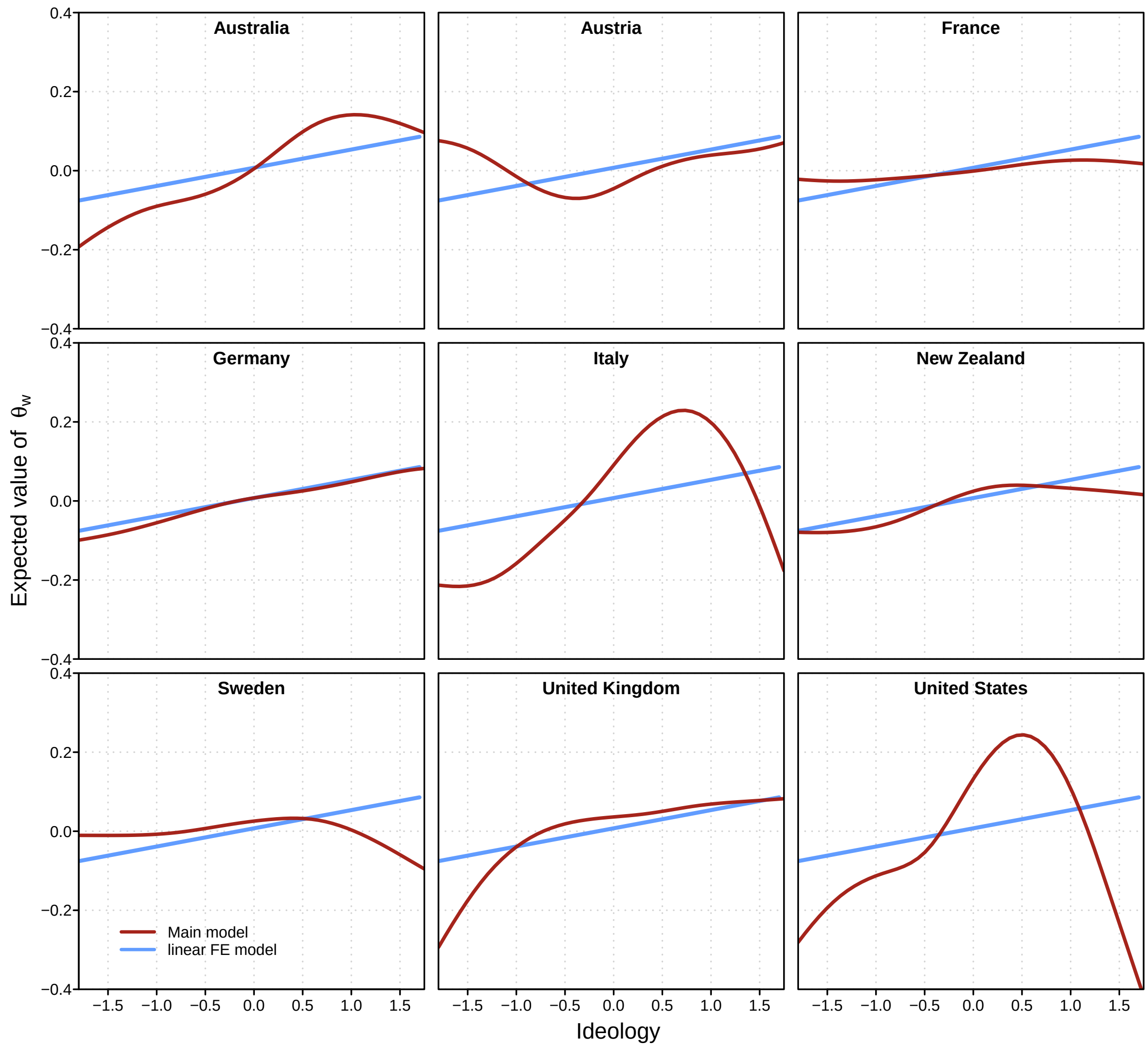

Figure D.5
Comparison of semiparametric and linear fixed effects models

This figure illustrates the difference between semiparametric estimates of ideology and a typical fixed effects linear model specification, which regresses θ_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 methology 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: Routeledge.

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.