

Approximate Measurement Invariance of Willingness to Sacrifice for the Environment Across 30 Countries: the Importance of Prior Distributions and Their Visualization

Arts, Ingrid; Fang, Qixiang; Schoot, Rens van de; Meitinger, Katharina

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Arts, I., Fang, Q., Schoot, R. v. d., & Meitinger, K. (2021). Approximate Measurement Invariance of Willingness to Sacrifice for the Environment Across 30 Countries: the Importance of Prior Distributions and Their Visualization. *Frontiers in Psychology*, 12, 1-18. <https://doi.org/10.3389/fpsyg.2021.624032>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>



Approximate Measurement Invariance of Willingness to Sacrifice for the Environment Across 30 Countries: The Importance of Prior Distributions and Their Visualization

Ingrid Arts*, Qixiang Fang, Rens van de Schoot and Katharina Meitinger

Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:

Alexander Robitzsch,
IPN—Leibniz Institute for Science and
Mathematics Education, Germany

Reviewed by:

Niki De Bondt,
University of Antwerp, Belgium
Artur Pokropek,
Institute of Philosophy and Sociology
(PAN), Poland

*Correspondence:

Ingrid Arts
i.j.m.arts@uu.nl

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 30 October 2020

Accepted: 21 June 2021

Published: 22 July 2021

Citation:

Arts I, Fang Q, van de Schoot R and
Meitinger K (2021) Approximate
Measurement Invariance of
Willingness to Sacrifice for the
Environment Across 30 Countries:
The Importance of Prior Distributions
and Their Visualization.
Front. Psychol. 12:624032.
doi: 10.3389/fpsyg.2021.624032

Nationwide opinions and international attitudes toward climate and environmental change are receiving increasing attention in both scientific and political communities. An often used way to measure these attitudes is by large-scale social surveys. However, the assumption for a valid country comparison, measurement invariance, is often not met, especially when a large number of countries are being compared. This makes a ranking of countries by the mean of a latent variable potentially unstable, and may lead to untrustworthy conclusions. Recently, more liberal approaches to assessing measurement invariance have been proposed, such as the alignment method in combination with Bayesian approximate measurement invariance. However, the effect of prior variances on the assessment procedure and substantive conclusions is often not well understood. In this article, we tested for measurement invariance of the latent variable “willingness to sacrifice for the environment” using Maximum Likelihood Multigroup Confirmatory Factor Analysis and Bayesian approximate measurement invariance, both with and without alignment optimization. For the Bayesian models, we used multiple priors to assess the impact on the rank order stability of countries. The results are visualized in such a way that the effect of different prior variances and models on group means and rankings becomes clear. We show that even when models appear to be a good fit to the data, there might still be an unwanted impact on the rank ordering of countries. From the results, we can conclude that people in Switzerland and South Korea are most motivated to sacrifice for the environment, while people in Latvia are less motivated to sacrifice for the environment.

Keywords: measurement invariance, visualization, Bayes, group ranking, MGCFA, prior sensitivity, Bayesian approximate measurement invariance (BAMI)

INTRODUCTION

One of the main issues the world population faces today is climate and environmental change. Some of the challenges that have to be faced include floods, droughts, food insecurity, and biodiversity loss. These challenges may give rise to socioeconomic problems such as refugee crises, relocating populations and cities, and famines (Zhang et al., 2020). As the challenges will differ across regions, but are not limited by national borders, international cooperation is required. At the same time,

a “one size fits all” solution is unlikely to solve these issues (Andonova and Coetzee, 2020). Several studies have been conducted on how the inhabitants of different countries perceive the subject of climate and environmental change, and the different aspects of social behavior regarding this subject: e.g., knowledge of climate change, risk perception, and the willingness to act (van Valkengoed and Steg, 2019). Hadler and Kraemer (2016) showed that the inhabitants of different countries do not assess all these threats in the same way: in some countries air pollution is seen as a major threat, while in others water shortages are considered a hazard.

The term “environmental concern” has been used widely to explain environmental behavior (e.g., Dunlap and Jones, 2002; Bamberg, 2003; Schultz et al., 2005; Franzen and Meyer, 2010; Marquart-Pyatt, 2012a; Fairbrother, 2013; Pampel, 2014; Mayerl, 2016; Pisano and Lubell, 2017; Shao et al., 2018). However, a clear definition of this concept is lacking (e.g., Dunlap and Jones, 2002; Schultz et al., 2005). Bamberg (2003, p. 21) described environmental concern as “the whole range of environmentally related perceptions, emotions, knowledge, attitudes, values, and behaviors,” while Dunlap and Jones (2002, p. 485) described environmental concern as “the degree to which people are aware of problems regarding the environment and support efforts to solve them and/or indicate the willingness to contribute personally to their solution.” Following the latter definition, environmental concern consists of at least two parts: on the one hand, perceptions of environmental problems (e.g., risks and beliefs), and, on the other hand, the willingness to contribute to the solution (e.g., to pay more taxes or higher prices, or to fly less). This translates into two latent variables that operationalize environmental concern: “environmental attitude” (EA) and “willingness to sacrifice (or pay) for the environment” (WTS). These two latent variables have been used both individually and in combination to operationalize environmental concern (Mayerl and Best, 2019). The latent variable WTS is frequently used to measure the extent to which people are willing to sacrifice something in their daily life (money, goods, time, comfort) to save the environment, and has been examined by several authors (e.g., Ivanova and Tranter, 2008; Fairbrother, 2013; Franzen and Vogl, 2013; Pampel, 2014; Sara and Nurit, 2014; Shao et al., 2018). The relation with cultural, sociological, economic, or political factors has been studied quite extensively (e.g., Marquart-Pyatt, 2012b; Franzen and Vogl, 2013; Pampel, 2014; Bozonnet, 2016; McCright et al., 2016; Shao et al., 2018).

Large-scale surveys are often used for exploring knowledge, attitudes, and (intentional) behavior regarding climate and environmental change (e.g., Bamberg, 2003; Franzen and Meyer, 2010; Marquart-Pyatt, 2012a; Hadler and Kraemer, 2016; Knight, 2016; Pisano and Lubell, 2017; Libarkin et al., 2018). One precondition for the valid comparison of attitudes toward climate and environmental change across many countries is that measurement properties are equivalent across countries (Jöreskog, 1971; Vandenberg and Lance, 2000). This means that all participants in all countries should interpret both the survey questions and the underlying latent variables in the same way. This equivalence of measurement properties is also called Measurement Invariance (MI). Establishing whether MI

holds is usually done by conducting a maximum-likelihood (ML) Multi-Group Confirmatory Factor Analysis (MG-CFA). There are at least four types of MI: configural (also referred to as “weak”), metric, scalar (“strong”), and residual (“strict”) invariance. Configural invariance allows for the comparison of latent variables among groups, metric invariance allows for a comparison of the items (questions) that make up the latent variable(s) among groups, and scalar invariance allows for the comparison of latent means across groups. Scalar invariance, however, is rarely established, especially when many groups are compared (e.g., Muthén and Asparouhov, 2013; Lommen et al., 2014; Kim et al., 2017; Marsh et al., 2018)¹.

Measurement invariance of the latent variable WTS has been investigated by Mayerl and Best (2019), and they established both configural and metric invariance, but not scalar invariance. Using ML MG-CFA, Marquart-Pyatt (2012b) also found configural and metric invariance, but not scalar invariance. To our knowledge, scalar invariance for the latent variable WTS has not been found by other authors, rendering the substantive interpretation of results from country rankings potentially untrustworthy (Byrne and van de Vijver, 2017; Marsh et al., 2018).

Alternative approaches have been proposed, such as alignment optimization, which allows for few but larger parameters differences between some groups (Asparouhov and Muthén, 2014), Bayesian Approximate MI (Muthén and Asparouhov, 2012; van de Schoot et al., 2013), hereinafter referred to as BAMI², which allows multiple but small differences between all groups, or a combination of both, BAMI alignment (Asparouhov and Muthén, 2014). When BAMI alignment is used, small variances are allowed for each group, while a few groups are allowed to have large variances. This leads to fewer noninvariant parameters than when the ML alignment method is applied, facilitating the interpretation of the model (Asparouhov and Muthén, 2014). Although this might be a highly interesting approach when a comparison of many groups is desired, it seems that, at least up until now, this approach has not been applied often: we only found two studies in which BAMI and alignment are combined: De Bondt and Van Petegem (2015) and van de Vijver et al. (2019), and certainly not in the field of environmental change.

The key to using Bayesian methods is the use of priors: some “wiggle room” is defined between which the variances of different groups are allowed to vary. However, the selection of these priors (from simulation studies, literature, or experience)

¹Residual invariance means that the sum of specific variance (variance of the item that is not shared with the factor and error variance) are also equal across groups (Putnick and Bornstein, 2016). Since this is not a requirement for comparing means across groups, we do not report it in this article.

²In previous research, the term Approximate MI (van de Schoot et al., 2013) has sometimes been used as a collective term for any method that can be used when the criteria for the exact scalar model are not fulfilled (Russell et al., 2016; Flake and McCoach, 2018), and sometimes to mention a specific method (e.g., Byrne and van de Vijver, 2017; Américo et al., 2020). To prevent any further confusion, we propose to use the term Bayesian Approximate Measurement Invariance (BAMI) when using a Bayesian model with strong informative priors on differences between factor loadings and/or intercepts, thus excluding non-Bayesian (ML or empirical Bayes) type of methods like random item effects (Fox and Verhagen, 2018).

is not an easy task. It seems that researchers applying Bayesian methods are not always fully aware of the potential impact of specifying priors (e.g., Spiegelhalter et al., 2000; Rupp et al., 2004; Ashby, 2006; Kruschke et al., 2012; Rietbergen et al., 2017; van de Schoot et al., 2017; König and van de Schoot, 2018; Smid et al., 2020). Nonetheless, for the verification and reproducibility of research (Munafò et al., 2017; van de Schoot et al., 2021), it is crucial to evaluate the influence of varying priors on the impact of substantive conclusions, which is referred to as sensitivity analysis. Some general guidelines regarding prior sensitivity can be found in the literature (e.g., Depaoli and van de Schoot, 2017; van Erp et al., 2018; van de Schoot et al., 2019; Pokropek et al., 2020). Although a sensitivity analysis of different prior settings helps to determine the impact of prior variances on substantive conclusions, it has, to our knowledge, never been applied for BAMl with empirical data.

The goals of our article are to apply the method of BAMl to the concept of “willingness to sacrifice (or pay) for the environment,” compare the results of different prior settings to each other and to other methods of dealing with measurement invariance (i.e., ML MGCFA and the ML alignment method) through visualization, and to provide an example for a transparent workflow.

In what follows, we first provide a technical introduction to the four methods we used to assess MI. As it can be difficult to interpret multiple models and methods, and because we want to be as transparent as possible in our decision-making process, we summarize our design choices and possible alternatives in a decision tree. We test the models to evaluate whether and how different prior variances influence the ranking of the countries on the latent variable WTS. We visualize the results to facilitate a comparison of the latent means of different models and methods without the use of complex and elaborate tables. All appendices, the scripts to reproduce our results, the final output files and additional material can be found on website of the Open Science Framework (OSF) (Arts et al., 2021).

TECHNICAL BACKGROUND

In this section, we introduce the four methods we used to evaluate measurement invariance: (1) ML MGCFA, (2) ML MGCFA using the alignment optimization, (3) BAMl, and (4) BAMl in combination with the alignment method.

MGCFA

The MGCFA model is defined as:

$$y_{ipg} = \nu_{pg} + \lambda_{pg}\eta_{ig} + \epsilon_{ipg} \quad (1)$$

where $p = 1, \dots, P$ is the number of observed indicator variables, $g = 1, \dots, G$ is the number of groups, $i = 1, \dots, N$ is the number of individual observations, λ_{pg} is a vector of factor loadings, ν_{pg} is a vector of intercepts and η_{ig} is a vector of latent variables. Furthermore, ϵ_{ipg} is a vector of error terms that is assumed to be normally distributed with $N(0, \theta_{pg})$, and η_{ig} is assumed to have a distribution of $N(\alpha_g, \varphi_g)$. θ_{pg} is the variance of ϵ_{ipg} , α_g is the mean of normally distributed latent variable η_{ig} , and φ_g is the variance

of η_{ig} . For WTS let $P = 3$ (3 items) and $G = 30$ (30 countries), which means that λ_{pg} is a 3×30 matrix. The same is true for ν_{pg} .

In the configural model, both λ and ν are allowed to vary across groups³, but the factor structure is equal for all groups, that is, in all 30 countries the latent variable WTS is covered by the same three items.

When both the number of latent variables and the factor loading λ are held equal across groups but the intercept ν is allowed to vary, one is testing for metric invariance: $\lambda_{11} = \lambda_{12} = \lambda_{13}$, etc. This means that for every group, the latent variable η_g contributes equally to item y_{pg} .

If metric invariance holds, it is possible to test for scalar invariance. In this case both loadings λ and intercepts ν are held equal across groups: $\lambda_{11} = \lambda_{12} = \lambda_{13}$ etc. and $\nu_{11} = \nu_{12} = \nu_{13}$ etc., so that Equation (1) becomes:

$$y_p = \nu_p + \lambda_p\eta + \epsilon_{ipg} \quad (2)$$

When scalar invariance holds, the latent means of WTS can be compared between groups, and a ranking of the latent means can be made. However, scalar, or strong, invariance is very rare, especially when comparing many groups (Asparouhov and Muthén, 2014; Byrne and van de Vijver, 2017; Kim et al., 2017; Marsh et al., 2018). This is due to the fact that with increasing number of countries, the probability increases that countries substantially deviate in answering behavior. When many groups with small deviations are being compared, these small deviations add up to the non-invariance of the scale assessing WTS.

Alignment Optimization

To reduce the impact of a lack of measurement invariance for many groups, the alignment optimization method has been introduced (Muthén and Asparouhov, 2013; Asparouhov and Muthén, 2014). Alignment optimization consists of two steps (Asparouhov and Muthén, 2014). First, a null model M_0 is estimated with loadings and intercepts allowed to vary across groups. As loadings and intercepts are freed across groups, factor means and factor variances are set to 0 and 1 for every group: $\alpha_g = 0$ and $\varphi_g = 1$. Now, the latent variable for the null model η_{g0} can be calculated.

Second, the method divides groups G into pairs Q and tries to find, for every Q , the intercepts and loadings that yield the same likelihood as the M_0 model (Asparouhov and Muthén, 2014; Flake and McCoach, 2018). Now, λ_{pg} and ν_{pg} can be calculated, where α_g and φ_g have to be chosen in such a way that they minimize the amount of measurement non-invariance and q_1 , q_2 , etc. are the different pairs of groups in the data. For the full set of equations, see Asparouhov and Muthén (2014), Flake and McCoach (2018). This means that, for the latent variable WTS, $q = 1 \dots 435$ for every item (for every item there are 435 possible pairs).

³Technically speaking, this is not entirely correct: for identification of the model, Mplus by default fixes the loading/intercept of the first item of every group to 1. For more details about parameterization of CFA models we refer the interested reader to Little et al. (2006).

The total amount of measurement non-invariance is shown by the total loss/simplicity function F :

$$F = \sum_P \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{pg_1, q_1} - \lambda_{pg_2, q_1}) + \sum_P \sum_{g_1 < g_2} w_{g_1, g_2} f(v_{pg_1, q_1} - v_{pg_2, q_1}) \quad (3)$$

In Equation (3), for the intercepts and loadings of every Q , the differences between the parameters are summed and then scaled by the Component Loss Function (CLF) f . Group sizes are appointed by weight factors w_{g_1} and w_{g_2} , where w_{g_1} is the weight factor of group 1 and w_{g_2} is the weight factor of the, differently sized group 2. In this way, bigger pairs of groups contribute more to the total loss function than smaller pairs. The weight factor can be calculated as follows:

$$w_q = w_{g_1, g_2} = \sqrt{N_{g_1} N_{g_2}} \quad (4)$$

The CLF has been used in exploratory factor analysis (EFA) to estimate factor loadings with the simplest possible structure (Jennrich, 2006). For the alignment the CLF is:

$$f(x) = \sqrt{\sqrt{x^2 + \epsilon}} \quad (5)$$

with ϵ being a small number, for example, 0.01 (Asparouhov and Muthén, 2014). This positive number ensures that $f(x)$ has a continuous first derivative, making the optimization of the total loss function F easier. As ϵ is so small, $f(x) \approx \sqrt{|x|}$, which leads to no loss if $x = 0$, amplified loss if $x < 1$, and attenuated loss if $x > 1$ (Asparouhov and Muthén, 2014). Due to this CLF, F will be minimized when there are a few large non-invariant loadings and intercepts and a majority of approximately non-invariant loadings and intercepts (Kim et al., 2017). When there are many medium-sized non-invariant parameters, the total loss function does not optimize (Flake and McCoach, 2018). If F does optimize, the parameters α_g and φ_g will be identified for all groups except the first one. For the first group, the variance can be calculated using the following parameter constraints, making the number of estimated parameters $(2G - 1)$:

$$\varphi_1 \times \dots \times \varphi_g = 1 \quad (6)$$

α_1 can be set to 0, although this is not always needed and might lead to untrustworthy estimates (Asparouhov and Muthén, 2014). When α_1 and φ_1 are both constrained, the alignment is called FIXED in Mplus, and when only φ_1 is constrained, that alignment is said to be FREE.

Although the alignment optimization allows for some large invariances between groups, all other groups are assumed to have the same loadings and intercepts. In other words, small variances between groups cannot be taken into account. To ensure that mean and variance can be fixed for one country (as it is in the MGCFAs and BAMI), we have to opt for the FIXED alignment and specify one country to be fixed.

BAMI

A synopsis of Bayesian statistics, including the most important aspects of determining prior distributions, likelihood functions and posterior distributions, in addition to discussing different applications of the method across disciplines can be found in van de Schoot et al. (2021).

With BAMI, priors with a mean of zero and some small variance are put on the differences between factor loadings and the differences between intercepts across groups: the terms λ_{pg} and v_{pg} from Equation (1) are now estimated being approximately equal across groups instead of exactly equal: $\lambda_{11} \approx \lambda_{12} \approx \lambda_{13}$ etc. instead of $\lambda_{11} = \lambda_{12} = \lambda_{13}$, etc. and $v_{11} \approx v_{12} \approx v_{13}$ etc. instead of $v_{11} = v_{12} = v_{13}$, etc.

The prior is not put directly on the differences between parameters, but on the covariances between parameters. This means that, for instance

$$V(\lambda_{11} - \lambda_{12}) = V(\lambda_{11}) + V(\lambda_{12}) - 2Cov(\lambda_{11}, \lambda_{12}) \quad (7)$$

where $V(\lambda_{11} - \lambda_{12})$ is the difference between the variances of the first loading of the first group and the first loading of the second group. If we assume that these prior variances are small, for instance 0.5, and the covariance is 0.495, that would lead to a value of 0.01 for $V(\lambda_{11} - \lambda_{12})$, or V^d .

BAMI uses strong informative priors on cross-group variances of loadings λ and intercepts v . It is important to carefully select these priors since they have a strong impact on the posterior results. Large values of V^d will result in decreasing the chance of model convergence, as they do not impose enough information on the model (Muthén and Asparouhov, 2012). Smaller values of V^d , on the other hand, might bring the model too close to a scalar model, reducing flexibility of the model to deal with the existing non-invariance.

BAMI With Alignment

The alignment method and BAMI can be combined. In that case, small variances are allowed for each group, while a few groups are allowed to have large variances. The alignment method for BAMI is similar to that for the exact method:

In the first step, an M_0 model is estimated, from which the optimal set of measurement parameters from the configural model is calculated. Now the M_0 model is a model where the intercepts and loadings are approximately equal across groups and the factor means and variances are estimated as free parameters in all groups but the first one.

In the second step, this M_{B0} model, the posterior of the configural factor loadings and intercepts are computed using the following equations:

$$\lambda_{pg,0} = \lambda_{pg,B} \sqrt{\varphi_{Bg}} \quad (8)$$

$$v_{pg,0} = v_{pg,B} + \alpha_{Bg} \lambda_{pg,B} \quad (9)$$

where $\lambda_{pg,0}$ and $v_{pg,0}$ are the configural loadings and intercepts and α_{Bg} , φ_{Bg} , $\lambda_{pg,B}$, and $v_{pg,B}$ are the BAMI parameters. Using

the BAMI parameters and Equations (8) and (9), the configural loadings and intercepts are computed for every iteration. These are then used to form the posterior distribution for $\lambda_{pg,0}$ and $\nu_{pg,0}$.

In the third and final step, the aligned estimates are obtained for every iteration using the configural factor loading and intercept values to minimize the simplicity function of Equation (3). The aligned parameter values obtained from one iteration are used as starting values in the next iteration. Finally, the aligned parameter values from all iterations are then used to estimate the aligned posterior distribution as well as the point estimates and the standard errors for the aligned parameters (Asparouhov and Muthén, 2014). This leads to fewer non-invariant parameters than when the ML alignment method is applied, facilitating the interpretation of the model (Asparouhov and Muthén, 2014).

METHODS AND DATA

Data

We used the data from the 2010 Module on Environment of the ISSP (ISSP Research Group, 2019). For the full report on this module, see GESIS (2019). The latent variable WTS consists of three questions, see **Table 1** for the exact wording, with answers on a five-point response scale (1 being *very unwilling* and 5 being *very willing*) and a *cannot choose* option for participants who could not or would not answer the question. WTS has, in combination with EA, been tested for MI by Mayerl and Best (2019) to explain the concept “environmental concern” when applied to 30 countries: Austria, Belgium, Bulgaria, Canada, Chile, Croatia, Czech Republic, Denmark, Finland, France, Germany, Great Britain, Israel, Japan, Latvia, Lithuania, Mexico, New Zealand, Norway, Philippines, Russia, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Turkey, and the United States. They found that, although metric invariance was achieved, scalar invariance was not. When we repeated this analysis we came to the same conclusion, for the results of this analysis, see Appendix A in Arts et al. (2021). For simplicity reasons, we only focus on the latent variable WTS, just like Ivanova and Tranter (2008), Fairbrother (2013), Franzen and Vogl (2013), Pampel (2014), Sara and Nurit (2014), and Shao et al. (2018). To further analyze this scale, we first ensured that we used the exact same data from the ISSP 2010 environment module and we followed the identical procedure as in the original study to handle missingness (i.e., listwise deletion—correspondence with author, November 26 2019), resulting in the same sample ($n = 24,583$). For the exact procedure and all code, see Appendix A in Arts et al. (2021). The sample sizes per country ranges from 798 (Iceland) to 3,112 (South Africa) with an average group size of 1,401, see for more details **Table 2**.

Analytical Strategy

We assessed the measurement invariance of the latent variable WTS by applying four methods for detecting MI: ML MGCEFA, the ML alignment optimization, BAMI, and BAMI with alignment optimization. For all analyses, one reference country was selected for which the factor mean and factor variance are held to 0 and 1, respectively (Spain). By fixing the mean and variance for a specific reference country for every model, it is

TABLE 1 | Exact wording of the questions in WTS.

Number	Question
Q12a	How willing would you be to pay much higher prices in order to protect the environment?
Q12b	How willing would you be to pay much higher taxes in order to protect the environment?
Q12c	How willing would you be to accept cuts in your standard of living in order to protect the environment?

TABLE 2 | Participating countries in the ISSP environmental module.

Country	Sample size	Country	Sample size
Argentina	1,130	Lithuania	1,023
Australia	1,946	Mexico	1,637
Austria	1,019	Netherlands	1,472
Belgium (Flanders)	1,142	New Zealand	1,172
Bulgaria	1,003	Norway	1,382
Canada	985	Philippines	1,200
Chile	1,436	Portugal	1,022
Croatia	1,210	Russia	1,619
Czech Republic	1,428	Slovakia	1,159
Denmark	1,305	Slovenia	1,082
Finland	1,211	South Africa	3,112
France	2,253	South Korea	1,576
Germany	1,407	Spain	2,560
Great Britain	928	Sweden	1,181
Iceland	798	Switzerland	1,212
Israel	1,216	Taiwan	2,209
Japan	1,307	Turkey	1,665
Latvia	1,000	United States	1,430
Total			50,437

ensured that any differences in outcomes are due to a method and not due to a difference in default settings of the model (for some models by default the parameters are fixed for the first group, while for other models it is the last group). We selected Spain as the reference country since the results presented by Mayerl and Best (2019) indicate that the results for WTS from this country can be seen as “average” within the group of thirty countries.

For the BAMI method, both with and without alignment, we tested the effect of different priors on the models. One way of selecting priors for new data is by using the results of simulation studies. **Table 3** shows an overview of simulation studies that have investigated BAMI and the priors that were used. As can be seen from this table, the simulation results are not entirely conclusive: The authors of these articles report that they achieve the best results when using priors with a variance of 0.001, 0.005, 0.01, or 0.05. However, the number of groups, group sizes and invariance criteria in these studies vary, complicating a comparison of the best performing prior variance(s).

We also searched for empirical studies in which BAMI was applied to empirical data. In a total of 30 empirical studies,

TABLE 3 | Simulation studies using Bayesian approximate measurement invariance.

Article	Number of groups	Group size	Prior variance	Invariance criteria
Muthén and Asparouhov, 2012	40	500	0.10, 0.05, 0.01	PPP
van de Schoot et al., 2013	2	1,000	0.50, 0.05, 0.01, 0.005, 0.0005	PPP, 95% CI
Kim et al., 2017	25, 50	50, 100, 1,000	0.05, 0.001	DIC, PPP, 95% CI, BIC
Lek et al., 2018	2	50, 100, 200, 1000	0.10, 0.05, 0.01, 0.001	95% CI
Shi et al., 2017	2	500	0.10, 0.05, 0.01	PPP, 95%CI
Pokropek et al., 2019	24	1500	0.10, 0.05, 0.01, 0.005	cor, RMSEA, 95%CI
Pokropek et al. (2020)	4, 24, 50	400, 1500, 3,000	0.05, 0.025, 0.01, 0.005, 0.001, 0.000*	BIC, DIC, PPP

*A Bayesian model with a prior variance of 0 is the scalar model.

PPP, posterior predictive p-value; DIC, deviance information criterion; 95% CI, 95% credibility interval; cor, correlation; RMSEA, root mean square error of approximation; BIC, Bayesian information criterion.

there were 13 in which only one prior was used, and in eight of these 13 studies, no specification was given as to why that specific prior was used. In the 17 studies where multiple priors were tested, three did not provide any information on why these priors were selected. The 14 other studies based the priors used on Muthén and Asparouhov (2012), van de Schoot et al. (2013), Asparouhov et al. (2015), or Seddig and Leitgöb (2018). For more information about these empirical studies and their variances see the additional material (Arts et al., 2021). The most frequently used prior variance in these studies is 0.01, followed by 0.05 as recommended by Muthén and Asparouhov (2012) and van de Schoot et al. (2013), respectively. However, other priors were also included in the different sensitivity analyses, ranging from 0.000000001 to 0.5.

We decided to estimate five different models, with priors with a variance of 0.05–0.01 (decreasing at 0.01 per prior) and three models with priors with variances of 0.001, 0.0005, and 0.0001. This includes the prior variances that are used most often in both simulation and empirical studies. Using such a large number of priors should create a clear overview of the influence of different prior variances on the rank order stability of the countries when ranked on their latent factor means. In addition to priors on the differences between loadings and intercepts, there are also priors on other parameters, such as the residuals. However, we will not discuss these priors in this article and we relied on the Mplus default values which can be found in Muthén and Muthén (2019). To ensure that the chains reached their target distributions, we checked whether all iterations after burn-in met the Gelman-Rubin criterion. Therefore, we set the convergence criterion to a rather strict 0.01 instead of the default 0.05 (Muthén and Muthén, 2019) and the

maximum and minimum number of iterations to 100,000 and 40,000, respectively.

For the analysis in this article, we used the software Mplus version 8.4 (Muthén and Muthén, 2019). The results were analyzed using R version 6.3.2 (R Development Core Team, 2017) and MplusAutomation version 0-7.3 (Hallquist and Wiley, 2018) was used for the exchange between the two programs. More information about the analysis and the exact Mplus and R code can be found in Appendix B on Arts et al. (2021).

Model Fit

To assess model fit for ML MGCFAs, the indices that are most widely used are the χ^2 -value, root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR) (Gallagher and Brown, 2013, p. 298). When testing for configural invariance cutoff values of CFI ≥ 0.95 , TLI ≥ 0.95 , RMSEA ≤ 0.06 , and SRMR ≤ 0.08 have been proposed by Hu and Bentler (1999). When checking for metric and scalar invariance, relative fit indices are more useful than absolute fit indices (Chen, 2007). These relative fit indices are a comparison of configural with metric and metric with scalar fit indices. Depending on these fit indices, the model for metric invariance can be assumed to perform better or worse than the model for configural invariance (and the same is true for metric and scalar invariance). For sample sizes above 300, Δ RMSEA ≤ 0.015 and Δ CFI ≤ 0.01 or Δ SRMR ≤ 0.03 indicate invariance when moving from the configural to the metric model, and Δ RMSEA ≤ 0.015 and Δ CFI ≤ 0.01 or Δ SRMR ≤ 0.01 indicate noninvariance when moving from the metric to the scalar model (Chen, 2007).

For the alignment method, fit indices have not been specified. Muthén and Asparouhov (2014) propose that the results can be considered trustworthy when no more than 25% of the parameters are non-invariant. However, Kim et al. (2017) have argued that this way the degree and location of non-invariance cannot be taken into account.

When BAMI is used, the model fit may be indicated by the posterior predictive p-value (PPP-value). This value indicates the ratio between the iterations for which the replicated χ^2 value exceeds the observed χ^2 value (Pokropek et al., 2020). A PPP-value of 0.50 indicates perfect model fit; a value below 0.50 indicates an underfit of the model, and a value above 0.50 indicates an overfit. Furthermore, the 95% credibility interval (CI) should include 0, preferably with 0 in the middle of the interval (Muthén and Asparouhov, 2012; van de Schoot et al., 2013). As PPP-values decline, the model fits the data less well. However, a specific cutoff value at which the model no longer fits the data is hard to determine. Muthén and Asparouhov (2012) suggest that models with PPP-values lower than 0.10, 0.05, or 0.01. do not fit the data anymore. In the literature, PPP-values above 0.05 are often seen as an indication for good model fit. A drawback of the PPP-value is that it might not identify a model with good fit correctly when using different priors with large sample sizes (Asparouhov and Muthén,

2010, 2019; Hoijtink and van de Schoot, 2018; Hoofs et al., 2018)⁴.

Recently, Bayesian versions of fit indices have been proposed: Bayesian RMSEA (BRMSEA), Bayesian CFI (BCFI), and Bayesian TLI (BTLI) can be computed based on differences between the observed and replicated discrepancy functions (Liang, 2020). These Bayesian fit statistics have been implemented in Mplus version 8.4, making it more convenient to identify good model fit (Asparouhov and Muthén, 2019). The calculation of these fit indices is very similar to that of the fit indices of an exact model, and therefore, the same cutoff values can be used (Asparouhov and Muthén, 2019; Garnier-Villarreal and Jorgensen, 2020). This means that $BCFI \geq 0.95$, $BTLI \geq 0.95$, and $BRMSEA \leq 0.06$ indicate good model fit. However, just as with the ML models, a combination of cutoff values must be used to indicate good or bad model fit. Other criteria that are being used to determine model fit are the BIC (Schwarz, 1978) and the DIC (Spiegelhalter et al., 2002). These information theoretic indices are less self-explanatory than the other fit indices: when selecting the best performing model from a series of models (the model that fits the data best and is the least complex), the model with the lowest BIC or DIC is preferred. This does not mean that the model with the lowest BIC or DIC is a good fit to the data: it is simply preferable to models with a higher BIC or DIC. Asparouhov et al. (2015) stated that, when sample sizes are large, coupled with a large number of observed indicators, DIC is preferable to BIC and Pokropek et al. (2020) concluded that DIC is a good indicator to identify the preferred prior mean and variance. On the other hand, Hoijtink and van de Schoot (2018) stated that the DIC is not suitable for evaluating models with small priors. This makes the use of the DIC as fit index promising, but also shows that its value should be treated with care. At a minimum, DIC should always be combined with other fit indices.

BAMI with alignment has, similar to the ML alignment method, no guidelines to determine model fit. Both De Bondt and Van Petegem (2015) and van de Vijver et al. (2019) tested a model with multiple small prior variances. De Bondt and Van Petegem (2015) used a prior variance of 0.01 and conducted a sensitivity analyses with prior variances decreasing with a factor 10, and van de Vijver et al. (2019) used a prior variance of 0.05 and conducted a sensitivity analyses with prior variances of 0.001,

0.005, 0.01, 0.05, and 0.1. Both De Bondt and Van Petegem (2015) and van de Vijver et al. (2019) analyzed the alignment part of the model by comparing, for each item, the intercepts and loadings across paired groups. This can be a very laborious process when multiple items and multiple groups are concerned. One could also use the rule of thumb that, to obtain trustworthy results, no more than 25% of the parameters can be invariant, as proposed by Muthen and Asparouhov (2014).

As shown above, there are many different criteria and cut-off values that provide insight into whether a model fits the data. Since there are so many different indicators these cutoff values should be treated with care: fit statistics can be influenced by, e.g., sample size or model complexity (Chen, 2007). Additionally, having one indication of good model fit is not enough to conclude that the model is a good fit to the data, and multiple fit statistics may even contradict each other. This exact point was addressed by Lai and Green (2016), who showed that RMSEA and CFI can contradict each other. Even when there is sufficient evidence that a model is a good fit to the data, this does not necessarily mean that it is the best model.

RESULTS

MGCFAs

The fit indices for the metric and scalar MGCFAs are shown in **Table 4**. Since the configural model was saturated, the results are not shown here. Therefore, for the metric model we assess the absolute fit indices instead of the relative fit indices. The metric model shows good fit, with a CFI and TLI of 0.993 and 0.989, respectively. With 0.069 the RMSEA value is above 0.06 but still below 0.08, indicating at least a reasonable fit. The fit indices for the scalar model all point to rejection of the scalar model: $\Delta RMSEA$, $\Delta SRMR$ and ΔCFI are well above the cutoff values of 0.015, 0.01, and 0.01 (0.085, 0.057, and 0.065, respectively). Based on these results we conclude that scalar invariance is absent, and that a comparison of the latent variable WTS across countries may not be trustworthy. However, this exact approach could be too strict in its assessment.

Alignment Optimization

Regarding the alignment optimization, the invariant, and non-invariant parameters are shown in **Table 5** with non-invariant parameters bolded and in brackets. Most non-invariant parameters can be found in the intercepts, with 48 non-invariant parameters, while for the loadings only seven parameters are non-invariant. However, a total of 55 parameters are non-invariant, which is 30.55% of all parameters. This is well above 25%, a rough cut-off value proposed by Muthen and Asparouhov (2014), implying that, for these data, a valid rank order comparison cannot be made if the ML alignment method is used.

BAMI

For BAMI, only the results for the models that converged are presented here (models with a prior variance of 0.02, 0.01, 0.001, 0.0005, and 0.0001). These models also converged when the

⁴Hoijtink and van de Schoot (2018) demonstrated that, with increasing sample sizes, the PPP-value does not decrease, but increases. Therefore, the prior-posterior predictive p -value (PPPP-value) was proposed by Hoijtink and van de Schoot (2018), and a generalized version was implemented in Mplus by Asparouhov and Muthén (2017). Whereas, the original PPP-value is a test of model fit which tests the fit of the model to the data and is based on comparing the model with the unrestricted covariance model, the PPPP-value is a test for the approximate parameters in the model. The PPPP-value is not a test for model fit and should not be interpreted as evidence that the model fits the data. The proper interpretation of the PPPP-value is given by Asparouhov and Muthén (2017): "If the test does not reject, the minor parameters (represented by θ_1) can be assumed to come from $N(0, \nu)$ distribution, with ν being a small variance. More broadly speaking, if the PPPP does not reject, that means that there is no evidence in the data for the minor parameters in model $M(\theta_1, \theta_2)$ to be outside the $N(0, \nu)$ distribution" (p. 10). Here, θ_2 represents the large parameters of model M . Unfortunately, the PPPP-value is not yet available for BAMI in Mplus.

TABLE 4 | Fit statistics of the MGCF model.

	χ^2 (df)	$\Delta\chi^2$ (Δ df)	p-value	RMSEA	Δ RMSEA	SRMR	Δ SRMR	CFI	Δ CFI	TLI	Δ TLI
Configural*											
Metric	287.324 (58)		0.00	0.069		0.051		0.993		0.989	
Scalar	2382.434 (116)	2095.110 (58)	0.00	0.154	0.085	0.108	0.057	0.928	0.065	0.944	0.045

*This model was saturated. df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; CFI, comparative fit index; TLI, Tucker-Lewis index. Numbers are absolute.

TABLE 5 | (Non)invariant parameters for ML alignment optimization.

Intercepts/Thresholds																														
Q12a	33	(40)	(56)	(100)	(124)	152	191	203	(208)	246	(250)	(276)	(376)	(392)	(410)	428	440	484												
		(554)	(578)	(608)	643	703	705	710	752	(756)	792	(826)	(840)																	
Q12b	33	40	(56)	(100)	124	(152)	(191)	(203)	208	246	250	276	376	392	410	(428)	(440)	(484)	554											
			(608)	(643)	(703)	705	(710)	(752)	756	(792)	(826)	(840)																		
Q12c	33	(40)	56	(100)	124	(152)	191	(203)	208	(246)	250	276	376	(392)	(410)	(428)	(440)	484												
		(554)	578	(608)	643	703	705	(710)	(752)	(756)	(792)	(826)	(840)																	
Loadings																														
Q12a	33	40	56	100	124	152	191	203	208	246	250	276	376	392	410	428	440	484	554	578	608	643	703	705	710	752	756	792	826	840
Q12b	33	40	56	100	124	152	191	(203)	208	246	250	(276)	376	392	410	428	440	484	554	(578)	608	643	703	705	(710)	752	756	792	826	840
Q12c	33	40	(56)	100	124	152	191	203	208	246	250	276	376	392	(410)	428	440	484	554	578	608	643	703	705	710	752	(756)	792	826	840

Noninvariant parameters are in bold and within parentheses.

number of iterations was doubled, which was not the case for the models with other prior settings.

To select the model(s) with a good fit, one could use model fit indices, but just as with regular SEM there is not one single statistic that should be used, and only a combination of fit indices should be used to indicate model fit. **Table 6** shows fit statistics for the models with prior variances 0.02, 0.01, 0.001, 0.0005, and 0.0001. **Table 6** shows that only for models with a prior variance of 0.02 and 0.01 the PPP > 0 (0.36 and 0.12, respectively) and the 95% CI contains 0. BRMSEA is 0.014 for the model with a prior variance of 0.02, and it is 0.049 for the model with prior variance of 0.01. For the other models, BRMSEA > 0.1. For the model with prior variance 0.02, both BCFI and BTLI are 1.00, and for the model with prior variance 0.01 BCFI is 0.999 and BTLI is 0.994, indicating good model fit. Using PPP-value, the model with variance 0.02 comes closest to 0.5 with a PPP-value of 0.36. However, the PPP-value might be untrustworthy because of the large sample size of our study (24,583 respondents) (Asparouhov and Muthén, 2019). Hoijtink and van de Schoot (2018) stated that the PPP-value is not suitable for evaluating small priors. Concerning both CI and BRMSEA, only the models with prior variances of 0.02 and 0.01 indicate a good fit. When looking at BCFI and BTLI, however, the models with prior variances of 0.02, 0.01, 0.001, and 0.005 all indicate good fit, although fit statistics approach to their cutoff values as prior variances decline. When combining the above results with the DIC for the different models, the values for the model with prior variance 0.02 is the lowest (19,7428.86), indicating that this is the best fitting model based on *post-hoc* fit indices.

Figure 1 shows the means of the latent variable for the BAMI models with variances of 0.02, 0.01, 0.001, 0.0005, and 0.0001.

From this figure it can be seen that with declining prior variance, the outcome of the model approaches that of the scalar model (on the right). This is to be expected, as the scalar model is a model of priors with a mean and variance of 0.

Figure 2 is a graph of the means per country per model (scalar invariance, ML alignment, and all BAMI models). For illustrative purposes, we present the results for BAMI both with and without alignment in one figure. **Figure 1** shows that the overall mean differences between latent means of the different models are small but increase as the prior variance decreases: $\Delta_{0.01-0.02}$ is 0.007 and $\Delta_{0.0005-0.0001}$ is 0.069. For individual countries, this is not always the case: **Figure 2** shows that, for the 15 lowest ranking countries this same pattern is visible, but for the top 15 countries the means increase with prior variance. However, as the countries rank lower, the differences between models increase. For the lowest-ranking country (Latvia) the difference between the model with prior variance 0.02 and that with prior variance 0.0001 is 0.616, while for the highest-ranking country (Switzerland) the difference is 0.208. For the three highest-ranking countries (Switzerland, South Korea and Denmark) the model with the highest prior variance (0.02) shows larger differences from the model with a prior variance of 0.01 (0.173, 0.161, and 0.128, respectively) than do other models with consecutively lower prior variances.

BAMI With Alignment

When the BAMI model with alignment is applied, first, the BAMI model is estimated. The outcome of the BAMI models is given in the description above (**Table 6**). From this BAMI model, a configural model is estimated, which is then aligned. This means that fit indices cannot be used to indicate model fit of the final

TABLE 6 | Fit statistics of the BAMI models.

Prior variance	PPP	95% CI	BRMSEA	BCFI	BTLI	BIC	DIC
0.02	0.363	[−52.706 to 79.836]	0.014	1.000	1.000	200288.68	197428.66
0.01	0.117	[−25.495 to 109.968]	0.049	0.999	0.994	200324.54	197514.18
0.001	0.000	[669.517 to 852.222]	0.117	0.976	0.968	201095.93	198186.01
0.0005	0.000	[1100.629 to 1286.848]	0.130	0.962	0.960	201546.82	198601.39
0.0001	0.000	[1873.721 to 2022.907]	0.148	0.938	0.949	202327.90	199334.78

PPP, posterior predictive probability; CI, credibility intervals; BRMSEA, Bayesian root mean square error of approximation; BCFI, Bayesian comparative fit index; BTLI, Bayesian Tucker-Lewis index; BIC, Bayesian information criterion; DIC, deviance information criterion.

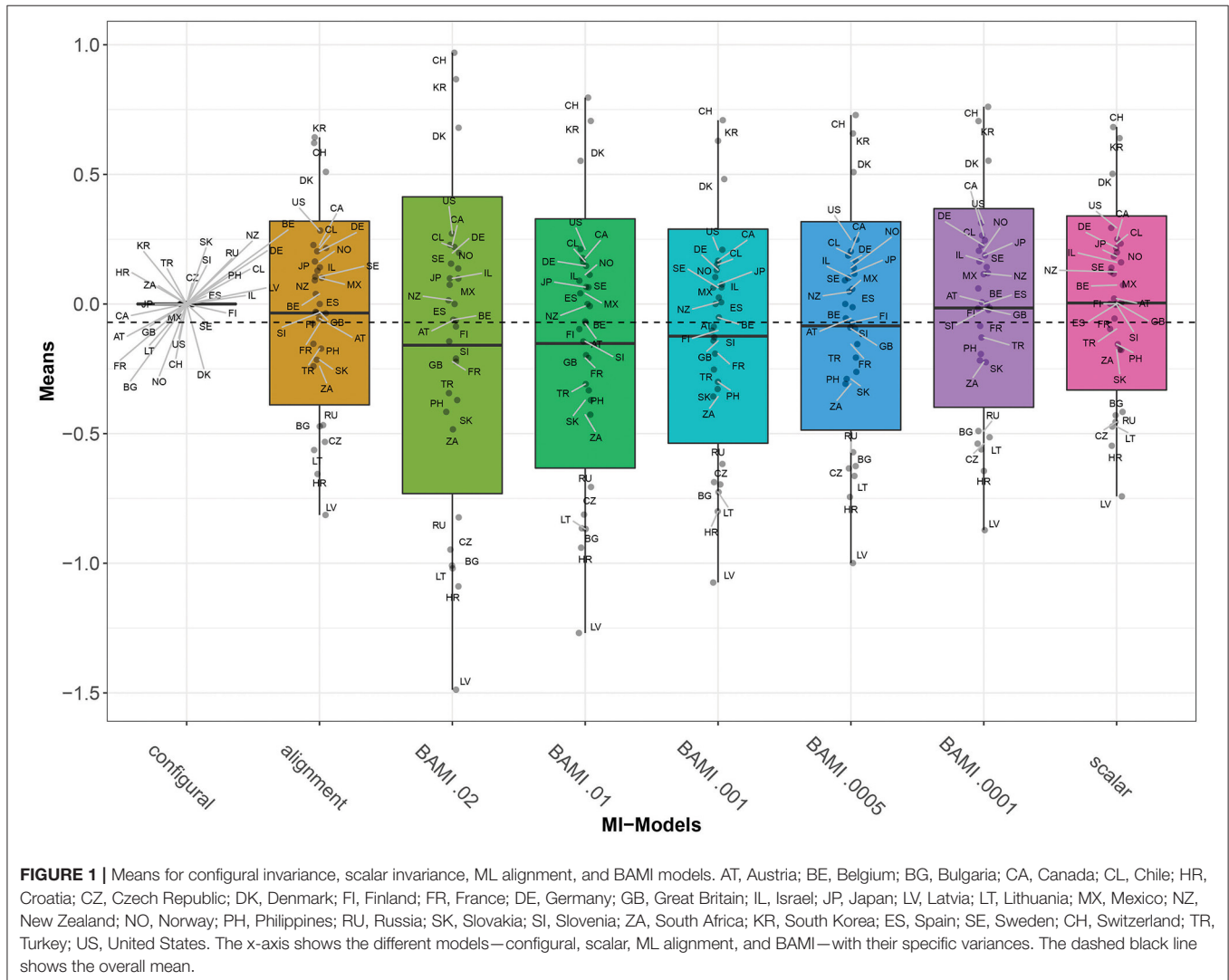


FIGURE 1 | Means for configural invariance, scalar invariance, ML alignment, and BAMI models. AT, Austria; BE, Belgium; BG, Bulgaria; CA, Canada; CL, Chile; HR, Croatia; CZ, Czech Republic; DK, Denmark; FI, Finland; FR, France; DE, Germany; GB, Great Britain; IL, Israel; JP, Japan; LV, Latvia; LT, Lithuania; MX, Mexico; NZ, New Zealand; NO, Norway; PH, Philippines; RU, Russia; SK, Slovakia; SI, Slovenia; ZA, South Africa; KR, South Korea; ES, Spain; SE, Sweden; CH, Switzerland; TR, Turkey; US, United States. The x-axis shows the different models—configural, scalar, ML alignment, and BAMI—with their specific variances. The dashed black line shows the overall mean.

model. Instead, just as with the ML alignment model, we use the percentage of non-invariant parameters to determine good model fit. **Table 7** shows the number of non-invariant parameters per model.

From this table, it can be seen that, as prior variances decrease, so does the number of non-invariant parameters. The three models with prior variances of 0.02, 0.01, and 0.001

all have a percentage of non-invariant parameters above 25% (although the model with prior variance 0.001 is only slightly above), making the results, and thus a group ranking from these models, unreliable. For the models with prior variances of 0.0005 and 0.0001 the percentages of non-invariant groups are 16.67 and 1.67, respectively, implying good model fit and a valid group ranking.

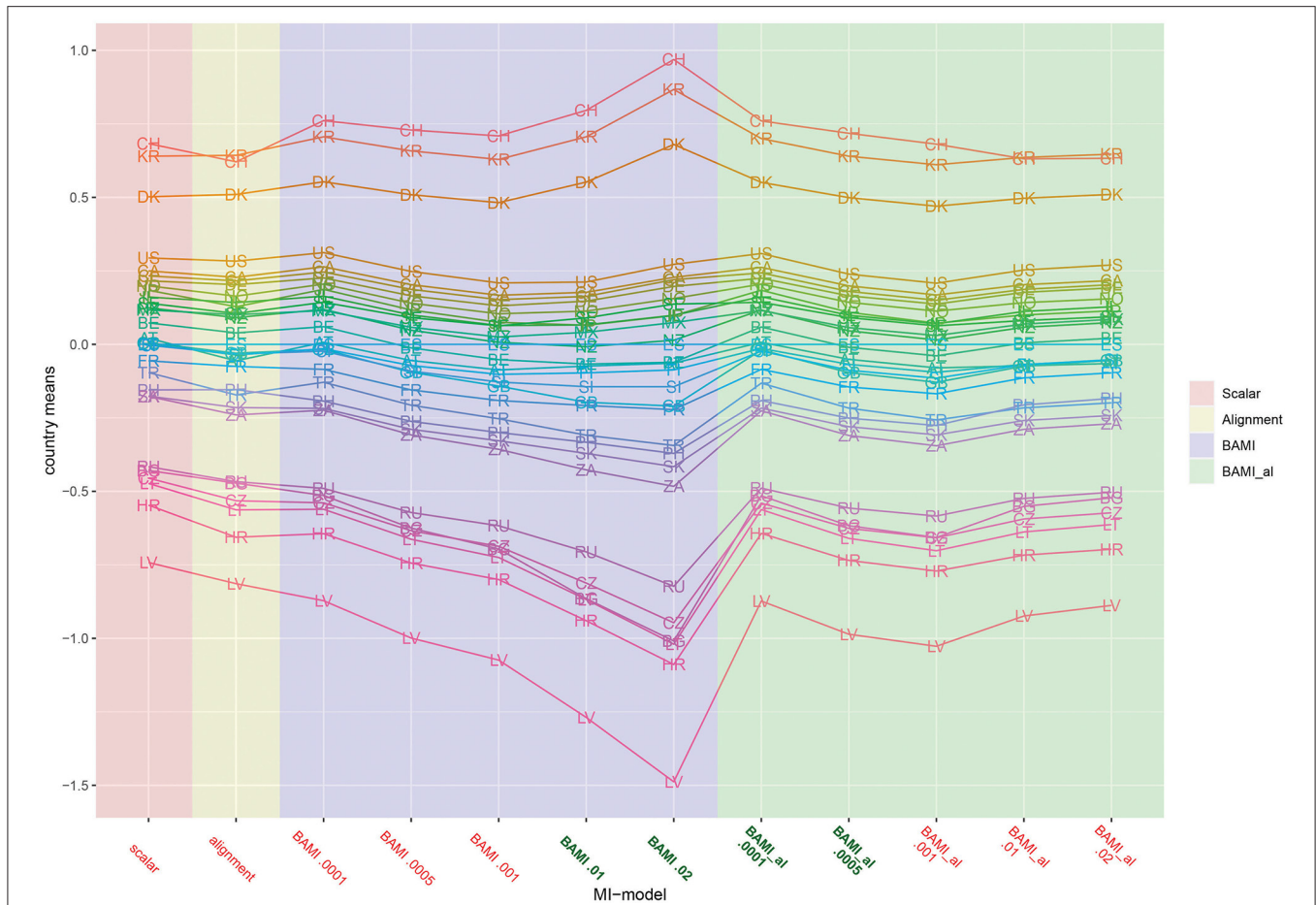


FIGURE 2 | Means per country for scalar invariance, ML alignment, and BAMI models with and without alignment. AT, Austria; BE, Belgium; BG, Bulgaria; CA, Canada; CL, Chile; HR, Croatia; CZ, Czech Republic; DK, Denmark; FI, Finland; FR, France; DE, Germany; GB, Great Britain; IL, Israel; JP, Japan; LV, Latvia; LT, Lithuania; MX, Mexico; NZ, New Zealand; NO, Norway; PH, Philippines; RU, Russia; SK, Slovakia; SI, Slovenia; ZA, South Africa; KR, South Korea; ES, Spain; SE, Sweden; CH, Switzerland; TR, Turkey; US, United States. The x-axis shows the different models—scalar, ML alignment, and BAMI with and without alignment—with their specific variances. Models that appear to be a good fit to the data are indicated in bold green, models with bad fit in red.

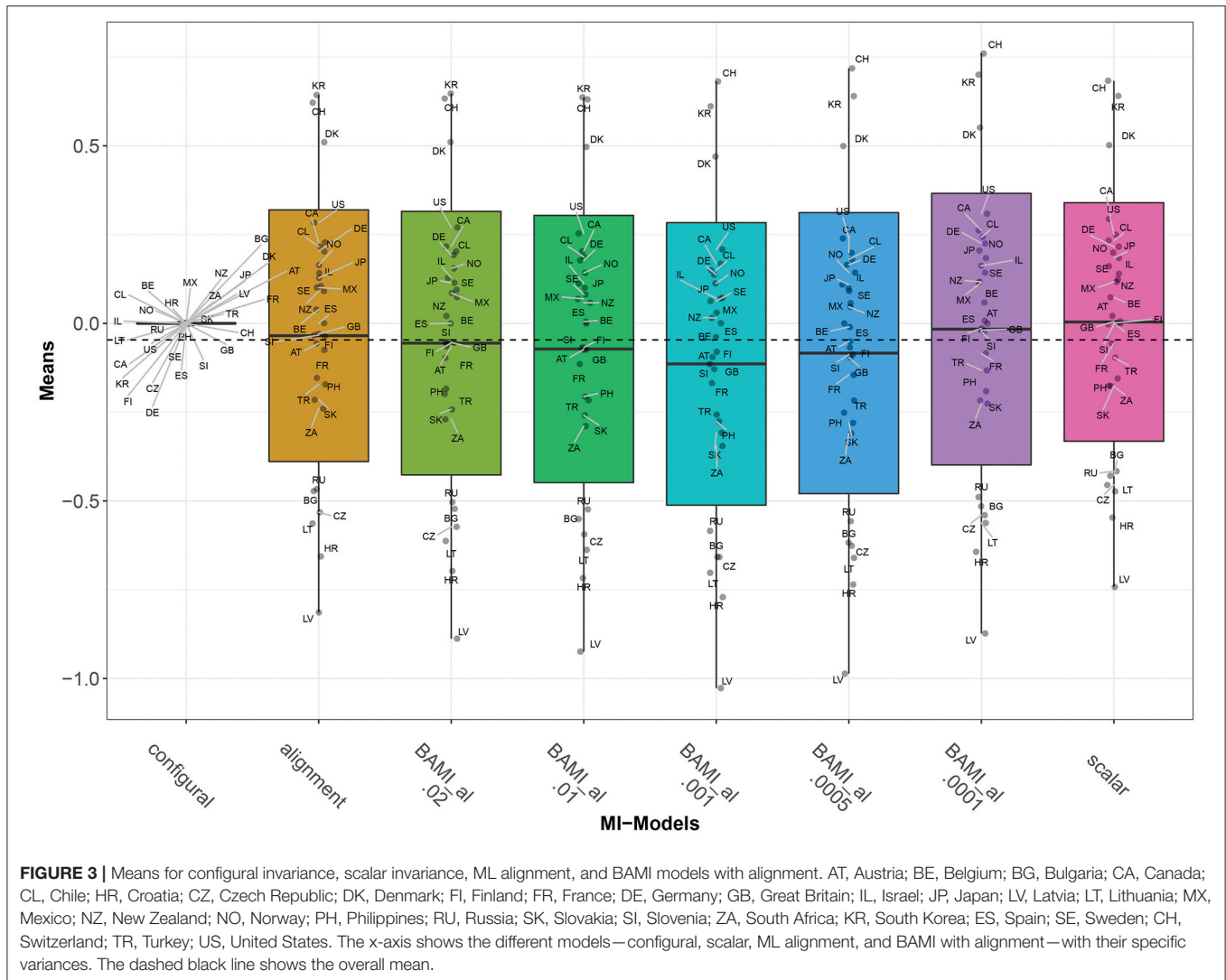
TABLE 7 | The number of non-invariant parameters for the BAMI models with alignment.

Prior Variance	Number of non-invariant intercepts			Number of non-invariant loadings			Total number	
	Q12a	Q12B	Q12C	Q12a	Q12B	Q12C	Sum	%
0.02	15	16	19	2	5	5	62	34.44
0.01	15	15	19	0	5	5	59	32.78
0.001	9	15	16	0	2	4	46	25.56
0.0005	4	11	12	0	1	2	30	16.67
0.0001	0	3	0	0	0	0	3	1.67

As with the ML alignment model, most non-invariant parameters are the intercept parameters. For the model with the lowest number of non-invariant parameters (prior variance 0.0001), these parameters belong to the intercepts of question 12b (are you willing to pay higher taxes to save the environment), for the countries Lithuania, South Africa, and Turkey. When taking into account only the models with a percentage of non-invariant parameters below 25%, there are 12 countries for which all

parameters are invariant for both models: Spain, Austria, Canada, Denmark, Finland, Germany, Israel, Mexico, Russia, Slovakia, Slovenia, and Sweden. **Figures 2, 3** show that, as the priors decrease, so do the mean differences of the model outcomes (both the overall means and the means per country).

Figure 3 shows that for the first three models with decreasing prior variance, the overall means also decrease. However, as prior variances decrease further (0.0005 and 0.0001), they rise



slowly toward the means of the scalar model. The differences between means of models with consecutive priors are less clear than for the BAMI models (Figure 1). Now, $\Delta_{0.01-0.02}$ 0.016, $\Delta_{0.001-0.01}$ 0.042, $\Delta_{0.0005-0.0001}$ 0.0675, and $\Delta_{0.001-0.0005}$ 0.031. Although this pattern is visible in the means per model (Figure 3), it is less distinctive when looking at the means of individual countries (Figure 2). In that case, this pattern is most pronounced for Latvia and, to a lesser extent, Bulgaria, Lithuania, Hungary, Russia, South Africa, Slovakia, Turkey, the Philippines, Denmark, Croatia, and Switzerland. Figures 2, 3 show that, as prior variances decrease, so do the mean differences of the model outcomes (both the overall means and the means per country). Again, Latvia is the country with the most pronounced differences when comparing different priors.

Ranking

Figures 1–3 show that the latent means vary depending on the choice of prior variance. Models with smaller prior variances seem to have outcomes that approach the outcome of the scalar model. However, there is some variation at the country level.

From Figure 2, we observe that there appear to be four different groups of countries with similar means: Switzerland, South Korea and Denmark at the top, then a large group with the United States, Canada, Chile, Germany, Norway, Japan, Israel, Sweden, New Zealand, Mexico, Austria, Great Britain, Finland, Spain, Slovenia, France, Turkey, Philippines, Slovakia, and South Africa. The third group comprises Russia, Czech Republic, Bulgaria, Lithuania, and Croatia, while the bottom group consists of only one country: Latvia. In particular for the second group, means are very close together, and it can be difficult to distinguish individual country means. Figure 4 shows the ranking of the 30 countries for the analyzed models that converged. This figure shows that for nearly all the models, ranking changes somewhat when a different prior variance is used. Upon closer inspection, 13 of the 30 countries occupy the same place in the ranking for all the models, and most changes appear to be in the middle of the ranking. When combined with Figure 2, it becomes clear that country mean differences are small, especially for the BAMI model with alignment. For the BAMI models, country means differ slightly more, especially at the top and the bottom of the

ranking. Mean differences for Latvia decrease with decreasing prior variance, and these differences are larger than the overall mean difference per model.

Comparing the individual country means of the BAMI and BAMI with alignment model shows that for all countries the differences between the models decrease with prior variance: differences between models are lowest when models with the lowest prior variances are compared. For the models with a prior variance of 0.0001 the country rankings are the same.

Focusing on only the models that appear to have a good fit to the data, according to their fit statistics, only the BAMI models with a prior variance of 0.02 and 0.01 and BAMI models with alignment with a prior variance of 0.0005 and 0.0001 are of importance. When comparing the rankings of these models, the ranking for the BAMI models is almost identical: only Spain and New Zealand switch places when changing models. The ranking of the BAMI models with alignment shows more variation: 23 countries rank the same for both models, while Mexico, New Zealand, Belgium, Austria, Great Britain, and Slovenia all shift one place up or down and Spain moves two places in the ranking.

These figures show that, regardless of prior variance or even model fit, people in Switzerland and South Korea are most motivated to sacrifice for the environment, while people in Bulgaria and Latvia are less motivated to sacrifice for the environment.

Decision Tree

As it can be difficult to draw conclusions from the means and rankings as shown in **Figures 1–4**, we devised a decision tree (**Figure 5**). This tree provides some insight into the decisions that we had to make regarding group means, group rankings and the influence of priors. Based on this decision tree, other readers might come to different conclusions. The tree comprises the entire process needed to evaluate the information contained in **Figures 1–4**, starting with the MGCFA test for scalar invariance:

1. We started with an ML MGCFA test for scalar invariance. To test for scalar invariance it is necessary that configural and metric invariance are met.
 - a. Yes: It is now possible to compare ranks.
 - b. No: Try another method to make means comparison valid. Go to step 2.
We did not find scalar invariance, so we followed the “no” arrow to step 2.
2. Do you expect a large difference in parameters for some groups and equality for the rest of the groups⁵?
 - a. Yes: Equality for almost all groups. Go to step 3.
 - b. No: Only small differences or small and large differences. Go to step 4.
We assumed there would be some differences in the parameters, although we did not know how large these

⁵When in doubt whether large difference in parameters for some groups are to be expected, it is advisable to consult a substantive expert of your field. The decision whether large parameters can be expected should be based on previous research and/or expertise.

differences would be or how many groups would differ from each other, so we first followed the “yes” arrow to step 3 and tested for ML alignment.

3. Does the alignment optimization yield < 25% non-invariant parameters?
 - a. Yes: The rank order can be trusted
 - b. No: Try another method to make means comparison valid. Go to step 4.
This yielded > 25% non-invariant parameters, so we followed the “no” arrow to step 4.
4. Do you expect only small differences in parameters for different groups⁶?
 - a. Yes: Only small differences. Use BAMI, Go to step 4A.
 - b. No: Both small and large differences. Use BAMI in combination with alignment optimization. Go to step 4B.
Now, we had to decide whether we expected small or large differences in parameters for different groups. Since we did not know how large the differences per group were, we used both, as they lead to step 5 in this decision tree; which option we chose would not make a difference.
5. Then, we needed to decide which priors to use and run the different models. We based our priors on previous literature on the use of BAMI (both simulations and empirical examples). We then moved to step 6.
6. We visualized the outcomes of the different models [scalar, ML alignment, and BAMI models (with and without alignment) that converged] in **Figures 1–4** (means and group rankings). The code to create these Figures can be found in Appendix C on Arts et al. (2021). We moved to step 7.
7. Is the rank order as a whole stable? (**Figure 4**)
 - a. Yes: No or only minor changes in rank. Then the rank order is not at all or only slightly influenced by the choice of priors.
 - b. No: Many changes across groups and models. Go to step 8.
Since there were numerous changes in the rank order, we did not consider the rank order stable and we followed the “no” arrow to step 8.
8. Does the pattern of the rank order across different models make sense?
 - a. Yes: Many changes, but all changes in the same section of the ranking (e.g., top) or the same groups change rank. Go to step 9.
 - b. No: The pattern seems erratic. Go to step 10.
From **Figure 5** we concluded that the upper and the lower parts of different rankings hardly change, and most rank changes take place in the middle part of the ranking. We considered this a logical pattern and followed the “yes” arrow.

⁶Similarly, when in doubt whether small differences for many groups are to be expected, it is advisable to consult a substantive expert of your field. The decision whether small parameter differences can be expected should be based on previous research and/or expertise.

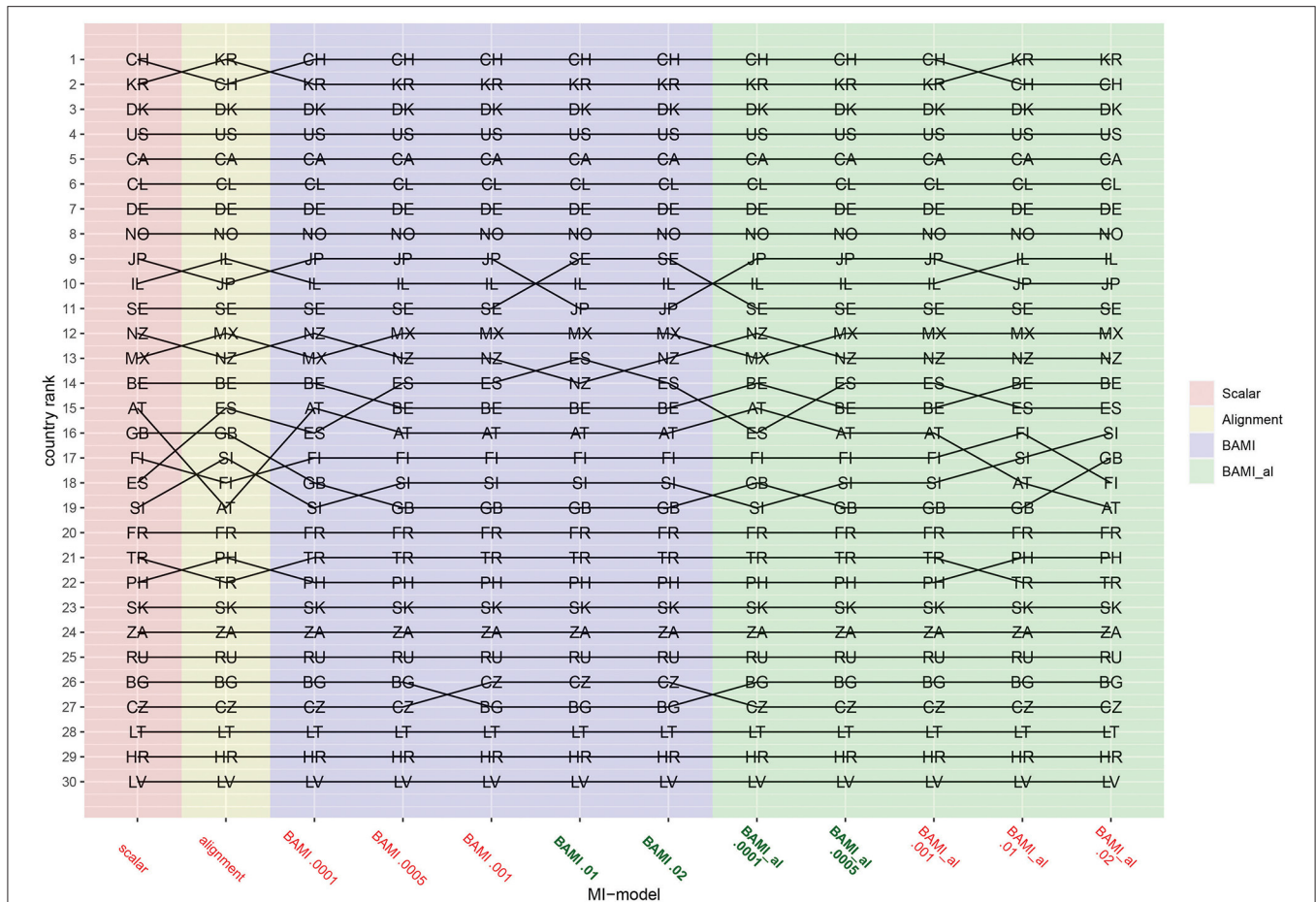


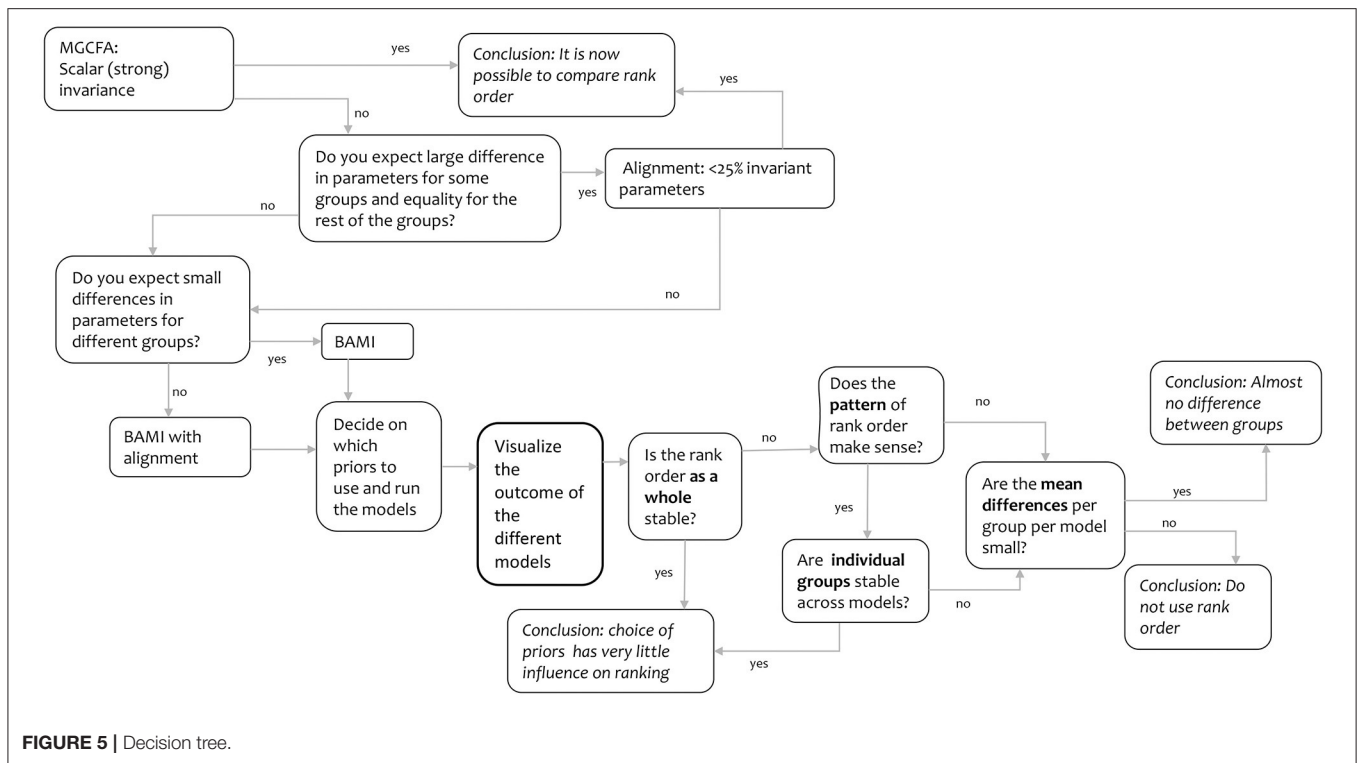
FIGURE 4 | Rankings per model. AT, Austria; BE, Belgium; BG, Bulgaria; CA, Canada; CL, Chile; HR, Croatia; CZ, Czech Republic; DK, Denmark; FI, Finland; FR, France; DE, Germany; GB, Great Britain; IL, Israel; JP, Japan; LV, Latvia; LT, Lithuania; MX, Mexico; NZ, New Zealand; NO, Norway; PH, Philippines; RU, Russia; SK, Slovakia; SI, Slovenia; ZA, South Africa; KR, South Korea; ES, Spain; SE, Sweden; CH, Switzerland; TR, Turkey; US, United States. The x-axis shows the different models—scalar, ML alignment, and BAMI with and without alignment—with their specific variances. Models that appear to be a good fit to the data are indicated in bold green, models with bad fit in red.

9. Are individual groups stable across models?
 - a. Yes: Individual groups never move more than one place up or down in the ranking across different models. Then the rank order is not or only slightly influenced by the choice of priors
 - b. No: Individual groups continue moving up or down the ranking across different models.
Changes in rank nearly always applied to the same countries, making the pattern rather stable. However, as some countries moved up or down two or three positions in the ranking across models, we found that stability of the groups could not be guaranteed. We followed the “no” arrow.
10. Are the mean differences per group per model small?
 - a. Yes. There is almost no difference between groups, and the influence of the priors is small.

- b. No: Do not use rank order.
Figure 3 shows that the differences per group are quite small, especially in the middle part of the ranking where most changes in rank take place. We therefore conclude that there is almost no difference between groups.

CONCLUSION AND DISCUSSION

The latent variable “willingness to sacrifice for the environment” (WTS) is an important aspect of environmental concern. It can provide insights into the intentional behavior regarding environmental concern, which, in turn, provides more insight into the willingness of the respondents to take action to protect the environment. Given that country rankings of latent means of WTS are frequently used in comparative studies, it is important to assess whether substantive findings are indeed trustworthy or are methodological artifacts due to lack of metric or scalar



invariance. The latent variable WTS was, in combination with the latent variable environmental attitude (EA), previously tested for MI by Mayerl and Best (2019). Using MGCFA, they did not find scalar invariance, questioning comparisons of the latent means across countries. However, recent discussions in MI point out that the approach of ML MGCFA may be too strict, and approaches such as alignment or BAMI, or a combination of both, may be a viable solution when exact scalar invariance tests fail (e.g., van de Schoot et al., 2013; Asparouhov and Muthén, 2014). In this article, we examined WTS in 30 different countries, using the 2010 ISSP data. We did not establish scalar invariance when using MGCFA, which is in line with the findings of Mayerl and Best (2019). In addition to MGCFA, we also assessed MI using ML alignment, BAMI and BAMI with alignment method.

Based on our results, we can determine which countries consistently rank high on the latent variable WTS (Switzerland and South Korea) and which countries consistently rank low (Latvia). However, we cannot say that, e.g., respondents in Sweden are more or less willing to sacrifice for the environment than respondents in Mexico. Thus, a more general conclusion about these country rankings can be drawn (high, low), but when exact ranking (e.g., fourth or fifth), or even exact means, are important, these country rankings should not be used. In conclusion, only with BAMI plus alignment optimization we were able to obtain stable results. From these, we can conclude that people in Switzerland and South Korea are most motivated to sacrifice for the environment, while people in Latvia are less motivated to sacrifice for the environment.

Regarding the use of different prior variances when using the BAMI method, models with a prior variance of 0.02 and

0.01 showed good model fit for most fit statistics (PPP, 95% CI, BRMSEA, BCFI, BTLI). For the model with variances of 0.001 and 0.0005 BCFI and BTLI were within limits. When taking into account that PPP might incorrectly identify model fit for models with large sample sizes (van de Schoot et al., 2012; Mulder, 2014), the results of BAMI models with a variance of 0.001 and 0.0005 might still fit the data. BIC and DIC are lowest for the BAMI model with a prior variance of 0.02, but the use of DIC for models with small prior variances has been disputed by Hoijtink and van de Schoot (2018). For the BAMI models with alignment the models with the smallest prior variances (0.0005 and 0.0001) give trustworthy results with a percentage of non-invariant parameters of 16.67 and 1.67%, respectively. This indicates that the BAMI models with a prior variance of 0.02 and 0.01 are a good fit to the data, while for the BAMI with alignment models, the models with a prior variance of 0.0001 and 0.0005 give trustworthy results.

Concerning comparing means of the BAMI models with different prior variances, both with and without alignment, the means are very similar for the models with prior variances of 0.0001 (difference of the overall mean per model is 0.001). These country rankings are, with the exception of Great Britain and Spain (rank 16 and 18, respectively) also the same for the scalar model. However, the scalar model and the BAMI model with a prior variance of 0.0001 cannot be assumed to be a good fit to the data (see **Tables 4, 6**), while the BAMI model with alignment with the same prior can. When comparing two models that indicate reliable outcomes—the BAMI with prior variance of 0.02 and the BAMI with alignment model with a prior variance of 0.0001—differences per country are much larger (ranging from 0.616 to

0.029, with the exception of 0 for Spain), thus indicating that prior variances can have a large influence on model outcomes, and that the model results that appear to be reliable, can be very close or even equal to the results of a model that should be rejected. This also shows the difficulty of comparing BAMI models to BAMI models with alignment: because the ground on which models should be rejected are very different, it is difficult to say which model should be preferred, if any.

It is our opinion that visualizing the results facilitates determining of the effect of different prior variances. A visual presentation of the results could be a valuable addition to the presentation of results in elaborated tables that can be challenging to interpret, especially when many groups are compared. The visualization approach that we used in this article is, however, not the only possibility to visualize (MI) results. Depending on e.g., research question, group size, and personal preferences, the researcher can choose other ways to visualize the results. For example, van de Schoot et al. (2015) chose to display the effect of different prior variances on the differences between groups in several line charts, while Pokropek et al. (2020) decided to use color-coded tables to identify the most suitable fit statistic to identify the optimal prior, Zercher et al. (2015) used scatter plots to represent latent means per country, and van de Vijver et al. (2019) used a 3-dimensional plot showing the Euclidean distances between different groups. However, most researchers still use (mainly) tables to present their results (e.g., Chiorri et al., 2014; De Bondt and Van Petegem, 2015; Gucciardi et al., 2016; Seddig and Leitgöb, 2018; Solstad et al., 2020; Vilar, 2020). We propose that a visual presentation of the results can improve the comprehension of test results, and can serve as a useful addition to previous presentations of results. This can be particularly useful when a researcher is faced with contradictory priors: a visualization of model outcomes with these different priors immediately shows the effect that the priors might or might not have. In particular researchers who are less familiar with the subject of Bayesian modeling might benefit from such a visual presentation.

Limitations

This study has some limitations, that need to be mentioned. First, the latent variable WTS is linked to intentional behavior, but intentional behavior alone cannot explain environmental concern. This one-scale, three-item model is an oversimplification of real-world data: multiple latent variables are required to provide insight into environmental concern. If MI holds for WTS in combination with other latent variables (e.g., EA) a country ranking would be more meaningful when determining nationwide environmental concern. Also, WTS is mainly financially driven (two out of three question refer to paying to protect the environment: see **Table 1**). This would mean that respondents who cannot or do not want to contribute financially but are willing to contribute in some other way affect this latent variable differently than those who are willing to sacrifice financially. Second, in the analysis of the different methods, most settings were Mplus default settings. Using different settings might lead to different outcomes: e.g., a different simplicity function when using alignment could affect

model outcome when the alignment optimization is used. It is also possible to choose Bayes as an estimator instead of ML. In that case, analysis starts with the same M_0 model as for ML alignment, and then loadings and intercepts are estimated using noninformative priors using Equation (3) (Asparouhov and Muthén, 2014). For other Mplus settings see Muthén and Muthén (2019). However, testing the many different settings in Mplus is beyond the scope of this article. Third, we used a decision tree (**Figure 5**) to interpret the results based on a visual inspection of country means and rankings. In this decision tree, several choices have to be made based on the ranking order and pattern (**Figure 4**). This pattern might not be interpreted by everyone in the same way: at Step 7 (is the rank order as a whole stable), Step 8 (does the pattern of the rank order make sense), and Step 9 (are individual groups stable across models) the reader has to decide whether a pattern is stable, the rank order makes sense, and individual groups are stable across models. It is also up to the reader to decide if the differences between group means are small or large (step 10). So, depending on the reader, conclusions might be different. However, we argue that using a decision tree always involves arbitrary decisions, like non-testable identification constraints, see for a discussion Little et al. (2006). We also believe that, in this case, the benefits of a decision tree (transparency to the workflow) outweigh the disadvantages.

Future Research

In this article we show that, for WTS, MI is present, making a ranking of countries possible. We also show that country means are not independent of specified priors and that, although the differences are small, an exact country ranking cannot be assumed. A combination of multiple latent variables (EA, knowledge of environmental concern, risk perception) might provide more insight into environmental concern. This would complicate model specification and analysis somewhat, since it would, e.g., make the use of priors on cross-loadings an important part of the model (Xiao et al., 2019; Liang et al., 2020). Another potential promising area of future study is the method of Robitzsch (2020), which improves the alignment optimization. A comparison with BAMI has not been made, yet. Future research may provide more insight into WTS and the topic of environmental concern. Looking at **Figure 2**, we see that four groups of countries have very similar means. It would be interesting to further investigate why this division into four groups appears. Is this also the case when other latent variables are investigated (separately or combined with WTS)? Is it purely data driven or are there underlying reasons that can explain these four groups (psychological, sociological, political, economic, etc.)? A multilevel model that includes such factors, might shed more light on why these four groups exist. The prior variances that we used in this article were based on previous literature on prior selection (both simulation and empirical studies). Another approach to selecting priors could be to consult literature on environmental change to determine the factors that drive environmental concern. Priors could then be based on e.g., socioeconomic status of a country, geography (and thus the environmental threat a particular

country faces) or political system and stability. Multilevel tests aiming to unravel these factors have been conducted by e.g., Marquart-Pyatt (2012a), Fairbrother (2013), Pampel (2014), and Pisano and Lubell (2017), but, to our knowledge, such factors have not yet been included in a Bayesian model. However, the use of such priors would emphasize the strength of Bayesian modeling.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: doi: 10.4232/1.13271. All appendices, the scripts to reproduce our results, the final output files and additional

material can be found on website of the Open Science Framework (OSF): <https://osf.io/mvyws/> (Arts et al., 2021).

AUTHOR CONTRIBUTIONS

IA, RvdS, and KM designed the study and wrote the article. IA conducted the analyses. QF provided the necessary code for visualization. All authors contributed to the article and approved the submitted version.

FUNDING

RvdS was supported by a grant from the Netherlands organization for scientific research: NWO-VIDI-452-14-006.

REFERENCES

- Amérigo, M., García, J. A., Pérez-López, R., Cassullo, G., Ramos, A., Kalyan-Venumbaka, S., et al. (2020). Analysis of the structure and factorial invariance of the Multidimensional Environmental Concern Scale (MECS). *Psicothema* 32, 275–283. doi: 10.7334/psicothema2019.281
- Andonova, L., and Coetzee, K. (2020). “Chapter 12: Does successful emissions reduction lie in the had of non-state rather than stat actors?,” in *Contemporary Climate Change, 1st Edn*, ed M. Hulme (New York, NY: Routledge), 177–190. doi: 10.4324/9780429446252-13
- Arts, I., Fang, Q., Meitinger, M., van de Schoot, R. (2021). *Approximate Measurement Invariance of Willingness to Sacrifice for the Environment Across 30 Countries: the Importance of Prior Distributions and Their Visualization*. OSF. doi: 10.17605/OSF.IO/MVYWS
- Ashby, D. (2006). Bayesian statistics in medicine: a 25 year review. *Stat. Med.* 25, 3589–3631. doi: 10.1002/sim.2672
- Asparouhov, T., and Muthén, B. (2010). *Bayesian Analysis of Latent Variables Models Using Mplus*. Mplus Web Notes, 1–60.
- Asparouhov, T., and Muthén, B. (2014). Multiple-group factor analysis alignment. *Struct. Equat. Model.* 21, 495–508. doi: 10.1080/10705511.2014.919210
- Asparouhov, T., and Muthén, B. (2017). *Prior-Posterior Predictive P-Values*. Mplus Web Notes, 1–16.
- Asparouhov, T., and Muthén, B. (2019). *Advances in Bayesian Model Fit Evaluation for Structural Equation Models*. Mplus Web Notes, 1–36. doi: 10.1080/10705511.2020.1764360
- Asparouhov, T., Muthén, B., and Morin, A. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: comments on Stromeier et al. *J. Manage.* 41, 1561–1577. doi: 10.1177/0149206315591075
- Bamberg, S. (2003). How does environmental concern influence specific environmentally related behaviors? A new answer to an old question. *J. Environ. Psychol.* 23, 21–32. doi: 10.1016/S0272-4944(02)00078-6
- Bozonnet, J. P. (2016). “Explaining environmental activism by national cultures: the hypothesis of hysteresis,” in *Green European: Environmental Behaviour and Attitudes in Europe in a Historical and Cross-Cultural Comparative Perspective* (New York, NY), 91–110.
- Byrne, B., and van de Vijver, F. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance. *Psicothema* 29, 539–551. doi: 10.7334/psicothema2017.178
- Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Model.* 14, 464–504. doi: 10.1080/10705510701301834
- Chiorri, C., Day, T., and Malmberg, L. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Front. Psychol.* 5:983. doi: 10.3389/fpsyg.2014.00983
- De Bondt, N., and Van Petegem, P. (2015). Psychometric evaluation of the overexcitability questionnaire-two applying Bayesian Structural Equation Modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Front. Psychol.* 6:1963. doi: 10.3389/fpsyg.2015.01963
- Depaoli, S., and van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: the WAMBS-checklist. *Psychol. Methods* 22, 240–261. doi: 10.1037/met0000065
- Dunlap, R. E., and Jones, R. E. (2002). “Chapter 15: Environmental concern: conceptual and measurement issues,” in *Handbook of Environmental Sociology, 1st Edn*, eds R. E. Dunlap and W. Michelson (Westport, WA: Abc-clio), 482–524.
- Fairbrother, M. (2013). Rich people, poor people, and environmental concern: evidence across nations and time. *Eur. Sociol. Rev.* 29, 910–922. doi: 10.1093/esr/jcs068
- Flake, J., and McCoach, D. (2018). An Investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Struct. Equat. Model.* 25, 56–70. doi: 10.1080/10705511.2017.1374187
- Fox, J. P., and Verhagen, J. (2018). “Random item effects modeling for cross-national survey data,” in *Cross-Cultural Analysis: Methods and Applications, 2nd Edn*, eds E. Davidov, P. Schmidt, J. Billiet, and B. Meuleman (New York, NY: Routledge), 529–550. doi: 10.4324/9781315537078-19
- Franzen, A., and Meyer, R. (2010). Environmental attitudes in cross-national perspective: a multilevel analysis of the ISSP 1993 and 2000. *Eur. Sociol. Rev.* 26, 219–234. doi: 10.1093/esr/jcp018
- Franzen, A., and Vogl, D. (2013). Acquiescence and the willingness to pay for environmental protection: a comparison of the ISSP, VVS, and EVS. *Soc. Sci. Q.* 94, 637–659. doi: 10.1111/j.1540-6237.2012.00903.x
- Gallagher, M. W., and Brown, T. A. (2013). “Introduction to confirmatory factor analysis and structural equation modeling,” in *Handbook of Quantitative Methods for Educational Research*, ed T. Teo (Rotterdam: Sense Publishers), 289–314. doi: 10.1007/978-94-6209-404-8
- Garnier-Villarreal, M., and Jorgensen, T. (2020). Adapting fit indices for bayesian structural equation modeling: comparison to maximum likelihood. *Psychol. Methods* 25, 46–70. doi: 10.1037/met0000224
- GESIS (2019). *ISSP 2010 Environment III, Variable Report: Documentation release 2019/06/13, related to the international dataset Archive-Study-No. ZA5500 Version 3.0.0, Variable Reports 2019|05*. Technical report, GESIS, Cologne.
- Gucciardi, D., Zhang, C.-Q., Ponnusamy, V., Si, G., and Stenling, A. (2016). Cross-cultural invariance of the mental toughness inventory among Australian, Chinese, and Malaysian athletes: a bayesian estimation approach. *J. Sport Rehabil.* 38, 187–202. doi: 10.1123/jsep.2015-0320
- Hadler, M., and Kraemer, K. (2016). “Chapter 1: The perception of environmental threats in a global and European perspective,” in *Green European: Environmental Behaviour and Attitudes in Europe in a Historical and Cross-Cultural Comparative Perspective*, eds A. Telesiene and M. Gross (New York, NY: Taylor & Francis), 13–30.
- Hallquist, M., and Wiley, J. (2018). MplusAutomation: an R package for facilitating large-scale latent variable analyses in Mplus. *Struct. Equat. Model.* 25, 621–638. doi: 10.1080/10705511.2017.1402334
- Hojtink, H., and van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive P-values. *Psychol. Methods* 23, 561–569. doi: 10.1037/met0000131

- Hoofs, H., van de Schoot, R., Jansen, N., and Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: simulation study introducing the BRMSEA. *Educ. Psychol. Measure.* 78, 537–568. doi: 10.1177/0013164417709314
- Hu, L., and Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118
- ISSP Research Group (2019). *International Social Survey Programme: Environment III - ISSP 2010*. ISSP Research Group.
- Ivanova, G., and Tranter, B. (2008). Paying for environmental protection in a cross-national perspective. *Austr. J. Polit. Sci.* 43, 169–188. doi: 10.1080/10361140802035705
- Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: the oblique case. *Psychometrika* 71, 173–191. doi: 10.1007/s11336-003-1136-B
- Jöreskog, K. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika* 36, 109–133. doi: 10.1007/BF02291393
- Kim, E. S., Cao, C., Wang, Y., and Nguyen, D. T. (2017). Measurement invariance testing with many groups: a comparison of five approaches. *Struct. Equat. Model.* 24, 524–544. doi: 10.1080/10705511.2017.1304822
- Knight, K. W. (2016). Public awareness and perception of climate change: a quantitative cross-national study. *Environ. Sociol.* 2, 101–113. doi: 10.1080/23251042.2015.1128055
- König, C., and van de Schoot, R. (2018). Bayesian statistics in educational research: a look at the current state of affairs. *Educ. Rev.* 70, 486–509. doi: 10.1080/00131911.2017.1350636
- Kruschke, J. K., Aguinis, H., and Joo, H. (2012). The time has come: bayesian methods for data analysis in the organizational sciences. *Organ. Res. Methods* 15, 722–752. doi: 10.1177/1094428112457829
- Lai, K., and Green, S. B. (2016). The problem with having two watches: assessment of fit when RMSEA and CFI disagree. *Multivar. Behav. Res.* 51, 220–239. doi: 10.1080/002737171.2015.1134306
- Lek, K., Oberski, D., Davidov, E., Ciecuch, J., Seddig, D., and Schmidt, P. (2018). “Approximate measurement invariance,” in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, eds T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, and B. Dorer, 911–929. doi: 10.1002/9781118884997.ch41
- Liang, X. (2020). Prior sensitivity in Bayesian structural equation modeling for sparse factor loading structures. *Educ. Psychol. Measure.* 80, 1025–1058. doi: 10.1177/0013164420906449
- Liang, X., Yang, Y., and Cao, C. (2020). The performance of ESEM and BSEM in structural equation models with ordinal indicators. *Struct. Equat. Model.* 27, 874–877. doi: 10.1080/10705511.2020.1716770
- Libarkin, J. C., Gold, A. U., Harris, S. E., McNeal, K. S., and Bowles, R. P. (2018). A new, valid measure of climate change understanding: associations with risk perception. *Clim. Change* 150, 403–416. doi: 10.1007/s10584-018-2279-y
- Little, T. D., Siegers, D. W., and Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Struct. Equat. Model.* 13, 59–72. doi: 10.1207/s15328007sem1301_3
- Lommen, M. J., van de Schoot, R., and Engelhard, I. M. (2014). The experience of traumatic events disrupts the measurement invariance of a posttraumatic stress scale. *Front. Psychol.* 5:1304. doi: 10.3389/fpsyg.2014.01304
- Marquart-Pyatt, S. T. (2012a). Contextual influences on environmental concerns cross-nationally: a multilevel investigation. *Soc. Sci. Res.* 41, 1085–1099. doi: 10.1016/j.ssresearch.2012.04.003
- Marquart-Pyatt, S. T. (2012b). Explaining environmental activism across countries. *Soc. Nat. Resour.* 25, 683–699. doi: 10.1080/08941920.2011.625073
- Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., et al. (2018). What to do when scalar invariance fails: the extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychol. Methods* 23, 524–545. doi: 10.1037/met000113
- Mayerl, J. (2016). “Chapter 8: Environmental concern in cross-national comparison: methodological threats and measurement equivalence,” in *Green European: Environmental Behaviour and Attitudes in Europe in a Historical and Cross-Cultural Comparative Perspective*, eds A. Telesiene and M. Gross (New York, NY: Taylor & Francis), 182–204.
- Mayerl, J., and Best, H. (2019). Attitudes and behavioral intentions to protect the environment: how consistent is the structure of environmental concern in cross-national comparison? *Int. J. Sociol.* 49, 27–52. doi: 10.1080/00207659.2018.1560980
- McCright, A. M., Dunlap, R. E., and Marquart-Pyatt, S. T. (2016). Political ideology and views about climate change in the European Union. *Environ. Polit.* 25, 338–358. doi: 10.1080/09644016.2015.1090371
- Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Comput. Stat. Data Anal.* 71, 448–463. doi: 10.1016/j.csda.2013.07.017
- Munafó, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie Du Sert, N., et al. (2017). A manifesto for reproducible science. *Nat. Hum. Behav.* 1, 1–9. doi: 10.1038/s41562-016-0021
- Muthén, B., and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol. Methods* 17, 313–335. doi: 10.1037/a0026802
- Muthén, B., and Asparouhov, T. (2013). *New Methods for the Study of Measurement Invariance With Many Groups*. Mplus Web Notes, 1–60.
- Muthén, B., and Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Front. Psychol.* 5:978. doi: 10.3389/fpsyg.2014.00978
- Muthén, L., and Muthén, B. (2019). *MPlus User Guide*.
- Pampel, F. C. (2014). The varied influence of SES on environmental concern. *Soc. Sci. Q.* 95, 57–75. doi: 10.1111/ssqu.12045
- Pisano, I., and Lubell, M. (2017). Environmental behavior in cross-national perspective: a multilevel analysis of 30 countries. *Environ. Behav.* 49, 31–58. doi: 10.1177/0013916515600494
- Pokropek, A., Davidov, E., and Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Struct. Equat. Model.* 26, 724–744. doi: 10.1080/10705511.2018.1561293
- Pokropek, A., Schmidt, P., and Davidov, E. (2020). Choosing priors in Bayesian measurement invariance modeling: a Monte Carlo simulation study. *Struct. Equat. Model.* 27, 750–764. doi: 10.1080/10705511.2019.1703708
- Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev.* 41, 71–90. doi: 10.1016/j.dr.2016.06.004
- R Development Core Team (2017). *R Version 6.3.2*. R Development Core Team.
- Rietbergen, C., Debray, T. P., Klugkist, I., Janssen, K. J., and Moons, K. G. (2017). Reporting of Bayesian analysis in epidemiologic research should become more transparent. *J. Clin. Epidemiol.* 86, 51–58.e2. doi: 10.1016/j.jclinepi.2017.04.008
- Robitzsch, A. (2020). LP loss functions in invariance alignment and haberman linking with few or many groups alexander. *Stats* 3, 246–283. doi: 10.3390/stats3030019
- Rupp, A. A., Dey, D. K., and Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to modeling. *Struct. Equat. Model.* 11, 424–451. doi: 10.1207/s15328007sem1103_7
- Russell, J. D., Graham, R. A., Neill, E. L., and Weems, C. F. (2016). Agreement in youth-parent perceptions of parenting behaviors: a case for testing measurement invariance in reporter discrepancy research. *J. Youth Adolesc.* 45, 2094–2107. doi: 10.1007/s10964-016-0495-1
- Sara, A., and Nurit, C. (2014). Pro-environmental behavior and its antecedents as a case of social and temporal dilemmas. *Br. J. Educ. Soc. Behav. Sci.* 4, 508–526. doi: 10.9734/BJESBS/2014/6573
- Schultz, P. W., Gouveia, V. V., Cameron, L. D., Tankha, G., Schmuck, P., and Franěk, M. (2005). Values and their relationship to environmental concern and conservation behavior. *J. Cross Cult. Psychol.* 36, 457–475. doi: 10.1177/0022022105275962
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136
- Seddig, D., and Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Survey Res. Methods* 12, 29–41. doi: 10.18148/srm/2018.v12i1.7210
- Shao, S., Tian, Z., and Fan, M. (2018). Do the rich have stronger willingness to pay for environmental protection? New evidence from a survey in China. *World Dev.* 105, 83–94. doi: 10.1016/j.worlddev.2017.12.033
- Shi, D., Song, H., Liao, X., Terry, R., and Snyder, L. A. (2017). Bayesian SEM for specification search problems in testing factorial invariance. *Multivariate Behav. Res.* 52, 430–444. doi: 10.1080/00273171.2017.1306432
- Smid, S. C., McNeish, D., Miočević, M., and van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small

- sample contexts: a systematic review. *Struct. Equat. Model.* 27, 131–161. doi: 10.1080/10705511.2019.1577140
- Solstad, B. E., Stenling, A., Ommundsen, Y., Wold, B., Heuzé, J. P., Sarrazin, P., et al. (2020). Initial psychometric testing of the coach-adapted version of the empowering and disempowering motivational climate questionnaire: a Bayesian approach. *J. Sports Sci.* 38, 626–643. doi: 10.1080/02640414.2020.1722575
- Spiegelhalter, D., Myles, J., Jones, D., and Abrams, K. (2000). *Bayesian Methods in Health Technology Assessment: A Review*. Technical Report, National Coordinating Centre for Health Technology Assessment (NCCHTA). doi: 10.3310/hta4380
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* 64, 583–639. doi: 10.1111/1467-9868.00353
- van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., and Van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *Eur. J. Psychotraumatol.* 6, 1–13. doi: 10.3402/ejpt.v6.25216
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., et al. (2021). Bayesian statistics and modelling. *Nat. Rev. Methods Primers.* 1:16. doi: 10.1038/s43586-021-00017-2
- van de Schoot, R., Hoijsink, H., Romeijn, J. W., and Brugman, D. (2012). A prior predictive loss function for the evaluation of inequality constrained hypotheses. *J. Math. Psychol.* 56, 13–23. doi: 10.1016/j.jmp.2011.10.001
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., and Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front. Psychol.* 4:770. doi: 10.3389/fpsyg.2013.00770
- van de Schoot, R., Veen, D., Smeets, L., Winter, S. D., and Depaoli, S. (2019). “Chapter 3: A tutorial on using the WAMBS,” in *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners, 1st edn*, eds R. van de Schoot and M. Miočević (New York, NY: Routledge), 30–49. doi: 10.4324/9780429273872-4
- van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: the last 25 years. *Psychol. Methods* 22, 217–239. doi: 10.1037/met0000100
- van de Vijver, F. J. R., Avvisati, F., Davidov, E., Eid, M., Fox, J.-P., Donné, N. L., et al. (2019). “Invariance analyses in large-scale studies,” in *OECD Education Working Papers* (Paris: Organisation for Economic Co-operation and Development (OECD)). doi: 10.1787/254738dd-en
- van Erp, S., Mulder, J., and Oberski, D. L. (2018). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychol. Methods* 23, 363–388. doi: 10.1037/met0000162
- van Valkengoed, A. M., and Steg, L. (2019). Meta-analyses of factors motivating climate change adaptation behaviour. *Nat. Clim. Change* 9, 158–163. doi: 10.1038/s41558-018-0371-y
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3, 4–69. doi: 10.1177/109442810031002
- Vilar, R. (2020). Basic Values Survey (BVS)-A 20-nation study. doi: 10.31234/osf.io/wvtj3
- Xiao, Y., Liu, H., and Hau, K. T. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Struct. Equat. Model.* 26, 665–677. doi: 10.1080/10705511.2018.1562928
- Zercher, F., Schmidt, P., Ciecuch, J., and Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: exact vs. approximate measurement invariance. *Front. Psychol.* 6:733. doi: 10.3389/fpsyg.2015.00733
- Zhang, D., Pei, Q., Frölich, C., and Ide, T. (2020). “Chapter 4: Does climate change drive violence, conflict and human migration?,” in *Contemporary Climate Change, 1st Edn*, ed M. Hulme (New York, NY: Routledge), 51–61. doi: 10.4324/9780429446252-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Arts, Fang, van de Schoot and Meitinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.