UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Jimenez Leal, W., Carmona, G., Murray, S., & Amaya, S. (2025). Validation of the Moral Foundation Vignettes in Latin America: The Scope of Moral Foundations Through the Lens of an Instrument. *Collabra: Psychology*, *11*(1). https://doi.org/10.1525/collabra.128178

Social Psychology

Validation of the Moral Foundation Vignettes in Latin America: The Scope of Moral Foundations Through the Lens of an Instrument

William Jimenez Leal[1,2a], Gino Carmona[1,2], Sam Murray[1,3], Santiago Amaya[1,4]

[1] Laboratorio de Juicios y Emociones y Morales, Universidad de Los Andes, Bogotá, Colombia, [2] Department of Psychology, Universidad de Los Andes, Bogotá, Colombia, [3] Philosophy Department, Providence College, Providence, RI, USA, [4] Department of Philosophy, Universidad de Los Andes, Bogotá, Colombia

Keywords: Moral Foundations Theory, loyalty, cross-cultural assessment, moral psychology, moral judgment

https://doi.org/10.1525/collabra.128178

Collabra: Psychology

Vol. 11, Issue 1, 2025

In this paper we examine the structural validity of the Spanish Translation of the Moral Foundations Vignettes, an instrument developed to measure moral judgement in the context of Moral Foundations Theory. With data from 3 countries (N = 1,650, through a polling agency) we identify a restricted set of items that fit the seven-factor solution implied by the theory. We conducted additional analyses (invariance testing and Differential Item Functioning) to examine the stability the results of the across countries. We found non-invariance and uniform differential functioning in a large number of items. Taken together these results suggest that although the current version of the MFV can be adequately used to measure moral judgement within societies, cross-cultural comparisons with this tool are restricted. Our validation raises questions about the cross-cultural validity of the instrument but also of some of the categories that underlie the intended measurements.

1. Introduction

Making fun of somebody with a disability, defying your boss, and urinating on someone's grave all seem morally wrong, albeit for different reasons. Poking fun at someone might cause emotional harm, whereas defying your boss is typically not a matter of harm. According to Moral Foundations Theory (MFT), moral judgment is sensitive to distinct domains of concerns that differentially inform judgments of wrongness. These distinct domains, or *foundations,* are characterized by their association with psychological adaptations that are culturally widespread and present in socially relevant normative judgment (Graham et al., 2013). The core formulation of MFT identifies five domains of moral concern: Care/Harm, Fairness/Reciprocity, Authority/Respect, Ingroup/Loyalty, and Purity/Sanctity (Graham et al., 2011, 2013). In turn, Graham et al. (2009) grouped these foundations into two superordinate categories: individualizing (Harm and Fairness) and binding foundations (Loyalty, Authority, Purity). Developments of the theory propose refinements of the foundations–such as distinguishing emotional from physical harm (Clifford et al., 2015)—as well as additional foundations, including liberty (Iyer et al., 2012), self-control (Hofmann et al., 2012), equality and proportionality (Atari et al., 2023). While MFT has been criticized on both conceptual (Gray & Keeney, 2015a, 2015b) and methodological grounds (Iurino & Saucier, 2020), the theory has also been used widely to explain moral judgment in several areas, from politics to education (for review, see Graham et al., 2018).

Several instruments have been developed to measure moral judgment within the framework of MFT. The most widely used is the Moral Foundations Questionnaire (Graham et al., 2009, 2011), validated in several languages and recently extended to include new items and divisions among existing foundations (Atari et al., 2023). This tool relies on ratings of highly abstract moral principles along the dimensions of relevance to moral decision making (e.g. "Whether or not someone's action showed love for his or her country" is relevant to decide if something is right or wrong) and moral judgment (e.g. agreement with statements such as "It is more important to be a team player than to express oneself"). Thus, responses to the MFQ create a snapshot of people's moral theories (Graham et al., 2009; Haidt, 2001) which might diverge from how people actually make moral judgments (Clifford et al., 2015). Other instruments that have been developed to capture moral judgment in a more concrete level of representation, rely-

a Correspondence should be addressed to William Jiménez Leal. Psychology Department. Universidad de los Andes. Cra 1 # 18-34, Bogotá, Colombia. Post Code 110311.

ing on images (Crone et al., 2018), Twitter posts (Hoover et al., 2020), or descriptions of concrete behaviors (Chadwick et al., 2006; Clifford et al., 2015).

The Moral Foundations Vignettes (MFV, Clifford et al., 2015) is a psychometric instrument designed to measure assessments of brief descriptions of concrete behaviors grouped according to MFT. The MFV was recently validated in Portuguese (Marques et al., 2020) and has the potential to allow rich cross-cultural comparisons that can be both theory driven and also empirically flexible to capture local variation (Barrett, 2020; Hruschka, 2020). It consists of 90 vignettes depicting moral violations across six foundations: Care (27 vignettes); Fairness (12); Authority (14); Loyalty (16); Purity (10) and Liberty (11) (a shortened version that exhibits good psychometric properties was recently developed; see Crone et al., 2021). All vignettes have the same structure: the respondent is asked to consider a third-party moral violation from their point of view (e.g. *You see* [a woman spanking her child with a spatula for getting bad grades in school]) and judge how morally wrong it is on a scale from 1 (not all wrong) to 7 (extremely wrong). The validated version of the MFV has been used to show relationships between purity violations and the intensity of moral judgment (Wagemans et al., 2018), partisan preferences and loyalty (Clifford, 2017), perceived social distance and moral valence (Dehghani et al., 2016). The MFV has also been used to test the primacy of care and fairness in the MFT (Isler et al., 2021).

Several challenges have been raised against MFT. Some concern the underlying pluralism of MFT (Beal, 2020; Fitouchi et al., 2023) and the adaptations underlying the evolution of distinct moral foundations (Curry et al., 2021). Other challenges concern the validity of the instruments developed within the framework of MFT. Generally speaking, validity challenges come in either of two varieties. One can challenge whether the instrument actually measures what the theory purports to explain. Also, one can challenge whether the instrument has adequate scope in light of the underlying theory (Borsboom, 2006).

Both challenges have been applied to the MFT and its corresponding measurement tools. On the first issue, for example, there are questions of whether the MFQ accurately captures the underlying complexity of moral judgments for different groups, given that its structure tends to vary significantly across samples. For example, Doğruyol et al. (2019) used secondary data from non-validated versions of the MFQ to show that the five-factor solution implied by the core version of MFT is stable for 30 WEIRD and non-WEIRD countries. However, Iurino and Saucier (2020) found weak evidence of measurement invariance for the (validated) short-form MFQ across 27 countries (spanning 5 continents). The source of these differences is debatable: it could be a matter of theory (e.g., foundations are represented differently in different cultural contexts) *or* it could be an issue with the MFQ as an instrument (e.g., the MFQ is highly sensitive and fails to represent the theory cross-culturally). On the second issue, instrument scope, there is no standard to evaluate whether different validated versions of an instrument are *equivalent* in terms of appropriately measuring the same construct (Iliescu, 2017). Does lack of equivalence suggest a limitation of the theory or a validation failure? Theory limitations can then be adjusted by using a bottom-up approach, where the use of an instrument suggest adjustments to the way the construct is conceived. For example, Atari et al. (2020) failed to establish the five factor solution for the MFQ with an Iranian sample. The results of their analysis suggested that loyalty was more central to overall moral concerns for this sample. Taken together, these data-driven results led the authors to propose a new moral foundation, only for Iran (Qeirat values), as opposed to the top-down approach favored by other proponents of MFT.

In consequence, although the main objective of the present study is to validate a Spanish version of the MFV, we believe this validation effort touches on several concerns about the relationship between theories of moral cognition and the instruments used to measure such theories. In addition to these overarching concerns about the relationship between methodology and theory, we also wanted to explore local characteristics of how moral judgment varies cross-culturally. We suspected that we would observe greater variability among vignettes within the Binding foundations (Loyalty, Authority, Purity) compared to Individualizing foundations for three reasons: (1) Clifford et al. (2015) explicitly selected loyalty violations on the basis of three requirements that result in a narrow characterization of loyalty compared to the original MFT; (2) Purity and Authority items on the MFQ exhibit greater variability in internal reliability compared to other foundations, which suggests that these foundations might be understood differently by participants in different contexts, and; (3) in the Portugese validation of the MFV (Marques et al., 2020), the largest number of vignettes discarded to maintain structural validity were, proportionally, from the purity and authority foundations.

Overview of the Current Studies

Following recent guidelines on scale development and validation (Clark & Watson, 2019; Flake et al., 2017), the current research is presented in three phases: Substantive (translation and initial item adjustments), Structural validation (factor analyses, reliability, [Studies 1 and 2]) and external validation (Convergent and discriminant validity, invariance testing [Study 3]). All participants were recruited through Netquest (polling agency). Research was approved by the IRB board of the Universidad de los Andes. Analyses were performed with R version 4.2.1 (R Core Team, 2022).

Data Availability

The data, materials, and supplementary information associated with this research are available at https://osf.io/mjtxw/.

2. Phase 1 (Substantive)

The full set of MFV items was translated into Spanish by two independent bilingual speakers (native speakers of

Colombian Spanish). Minor differences in translations (mainly the use of equivalent nouns) were discussed and solved by the translators. Next, two of the authors reviewed each item and proposed culturally adapted alternatives to Colombian Spanish, when necessary, keeping the meaning and the magnitude of the offense as close as possible to the original items. This adapted set of items was back translated by an independent translator and then passed onto the other two authors, who evaluated whether the original sense of the items was preserved. For this step, we took into account modifications suggested by Marques et al (2020) in order make easier any possible comparisons between Brazil and other Latin-American countries. This last round of revisions resulted in the final Colombian Spanish version used in Study 1, which was submitted to two native speakers of Peruvian and Argentinian Spanish to adjust any other cultural differences that could improve cultural appropriateness of the items for these countries.

The final set of items items used in all countries can be found at the OSF page for the project. Most vocabulary changes centered on the object of the transgression. For example, the item *You see a US swimmer cheering as a Chinese foe beats his teammate to win the gold* was translated into Colombian Spanish as *You see a Colombian cyclist celebrating as his teammate is beaten by a Chinese opponent* given the salience of cycling in Colombia as a sport where individual athletes are identifiable. Following the same logic, we used surfing in Peru and tennis in Argentina. These changes made the transgression more natural to the local variant of Spanish and did not change the intended meaning.

More significant changes were introduced in two of the loyalty items (*You see a coach celebrating with the opposing team's players who just won the game* and *You see a former US General saying publicly he would never buy any American product*). The former was deemed too abstract in the cultural adjustment review, and thus was replaced by introducing the names of actual soccer teams for each country. The latter was considered an unrealistic moral transgression by experts and it was replaced by an explicit negative comparison of a local product by a relevant political public figure. Average number of character and words did not differ significantly from the original version (US: average of 15 words and range 13-18, average of 83 characters and range of 69-91, Spanish: average of 17 words and range of 10-25 words, average of 89 characters, and range of 60-130).

3. Phase 2. Structural Validation

3.1. Study 1

The objective of this study is to examine the structural validity of the MFV with a Colombian sample as proposed by Clifford et al (2015), which established a six-factor solution according to the MFT taxonomy (care, fairness, authority, purity, loyalty and liberty).

3.1.1. Participants

618 people took part in this study (52% female, 48% male) with ages ranging between 18 and 76 (M = 41.30, SD = 14.54) and contacted through a polling service. 32% of the sample reported having a bachelor's degree, 27% reported having a technical degree, 18% reported having achieved a high school diploma, 16% reported having a master's degree, and less than 1% (2 participants) reported having doctorates.

3.1.2. Materials and Procedure

All participants completed the following self-reported measures.

Moral Foundation Vignettes. We used the 90 vignettes recommended by Clifford et al. (2015) plus 5 neutral scenarios also used in the original study as controls. For each vignette, participants had to judge how morally wrong was the action of the person [*¿Qué tan moralmente incorrecta considera que fue la acción de la persona?*] using a scale from 1 (not wrong at all [*nada mala en absoluto*]) to 5 (extremely wrong [*extremadamente mala*]). Immediately after, participants had to report why they thought the action was morally wrong [*¿Por qué cree que esta acción estuvo mal?*] and response choices corresponded with each of the moral foundations (e.g. it violates norms of purity (e.g., degrading or disgusting acts) [*Viola las normas de pureza (por ejemplo, actos degradantes o repugnantes)*]). The format for both questions corresponds exactly to Clifford et al. (2015).

Sociodemographic characteristics. Participants reported their age, gender, education level, religion and political preference. The latter was asked with the question *Which of the following best describes your overall political preference?* With responses ranging from 1 (Strongly liberal) to 6 (Strongly conservative).

The survey was administered through Qualtrics. Item presentation was fully randomized for the MFV, which was always followed by questions on sociodemographic information, in a fixed order.

3.1.3. Results

Table 1 presents the summary of classification and average wrongness judgement by item and foundation. Overall average wrongness was 3.46 (SD = 0.66) and the average successful classification rate (e.g. classifying an infraction under the intended foundation) was 64%. By foundation, loyalty violations had the lowest perceived wrongness (M = 3.33, SD = .32) and also the lowest classification success rate (57%) in contrast with justice and liberty violations, that had the highest average perceived wrongness (M =

3.82, *SD* = .54) and classification rate (74%), respectively. Additionally, three vignettes (two of the purity foundation[1] and one from the care foundation[2]) were classified as "not being morally wrong" at least 50% of the time. Overall, 29 of the 90 original items did not meet the inclusion criteria set in the original development of the scale (60% classification accuracy and less than 20% classification in other categories).

We proceeded to run exploratory factor analyses with and without the problematic items. KMO values ranged between .8 and .96 (overall of .95) and the items were strongly correlated (Bartlett test $\chi 2$ = 26782, p < 0.001). Parallel analyses suggested five factors while the original test comprised six factors and a recent validation (Marques et al, 2021) found seven factors (splitting care into physical and emotional), so we fitted rotated (Promax) and non-rotated solutions with five, six and seven factors. We also fitted an additional rotated eight-factor solution to discriminate between animal and human care but did not consider it further as it exhibited poor fit (See supplementary material).

These analyses suggested retaining the seven-factor rotated solution with the full data set. This solution exhibited good fit indices (*RMSEA* = 0.028 [0.027, 0.03], *TLI* = .90). The full factor loadings table is presented in the supplemental materials, where only factor loadings greater than .35 are displayed. Items for liberty, loyalty and purity clearly load together (factors 3,4 and 5). The other factors present a mix of themes. Factor 1 contains mostly care items, but also includes four fairness items and items from purity and loyalty. Factor 2 comprises mostly items from the authority foundation but also includes several fairness items. Factors 6 and 7 refer exclusively to emotional care and physical care, respectively. Overall, fairness items are interspersed among several factors and factor loadings tend to be around .4, indicating a good load. There are however some vignettes that failed to adequately load onto any factor (class, brother, judge, date, attractive, disfigured). All of these also failed to display adequate classification rates (except *class* from the authority foundation).

We ran confirmatory factor analyses comparing the six- and seven-factor solutions (see Table 2). The best models are the ones with seven factors (where care is split into emotional and physical) and the model with the best fit is the seven-factor solution fitted to the items retained in the prior process, likely because of estimating fewer parameters. We also included alternative models including modification indices where covariances were estimated for items within foundations, but these did not significantly improve the overall fit and are not reported here.

The final set of items exhibit excellent internal consistency: Care (α =.92, ω = .93), Emotional Care (α =.85, ω =.88), Physical Care (α =.92, ω = .9), Fairness (α =.84, ω =.87), Authority (α =.87, ω =.88), Loyalty (α = .81, ω = .83), Purity (α = .76, ω = .82) and Liberty (α = .88, ω = .90). With this set of items, we estimated correlations between scores derived from the average of each foundation per participant and reported political ideology (highest numbers indicate conservatism). The highest correlations were with purity (.13), authority (.10) and loyalty (.11), compared with Care (.03), Fairness (.08) and Liberty (-.03), all significant at p < 0.001. These correlations are similar in direction to those found by Clifford et al (2015) but overall smaller in magnitude.

3.1.4. Discussion

The final set of items validated items includes 60 vignettes: Care (18), Justice (8), Liberty (10), Authority (10), Purity (6) and Loyalty (8). This is similar to the validated Brazilian Portuguese version (68 items) but somewhat shorter than the original US version (90 items). The items in the Purity and Loyalty foundations are particularly problematic, where 40% and 50% of the items were discarded. While we are confident that this is a tool that can be used to adequately probe moral judgment with Colombian participants and to make meaningful comparisons with other validated versions, there are reasons to doubt the generality of the Purity and Loyalty vignettes (Murray et al., 2024). This pattern could be completely idiosyncratic to Colombian participants and not generalize to Spanish speakers from other countries. In Study 2 we pursue this question while also including alternative measures of moral judgment to assess the convergent validity of the MFV.

3.2. Study 2

The objective of this Study was to further examine the structural validity of the MFV with a larger sample of participants from Perú, Argentina and Colombia. It also assesses convergent validity by examining how the MFV is associated with other measures of moral judgment, the Morality as Cooperation Questionnaire (MAC-Q) and Moral Foundations Questionnaire (MFQ).

3.2.1. Participants

1033 participants were recruited from Peru (400, 48% female, 52% male), Argentina (400, 49% female, 51% male) and Colombia (233, 49% female, 51% male) using the same polling firm. None of the participants took part in Study 1 and data from one participant from Peru was excluded for incomplete responding. Colombian participants were contacted in two waves. 73 participants had to be contacted again due to a coding error. Education levels were similar to participants in Study 1, with most participants reporting having at least a college degree (26% Colombia, 24% Peru

1 "*You see a single man ordering an inflatable sex doll that looks like his secretary*" and "*You see two first cousins getting married to each other in an elaborate wedding.*"

2 "*You see a girl telling a boy that his older brother is much more attractive than him*".

Table 1. Classification rate and mean wrongness ratings by vignette and foundation for Study 1

Items	Care	Fairness	Liberty	Authority	Purity	Loyalty	Not Wrong	Mean Wrongness
amp	**71.3%**	3.9%	4.5%	5.8%	12.0%	1.1%	1.3%	4.46
play	**62.2%**	1.8%	7.6%	6.3%	6.7%	7.0%	8.4%	3.51
fat	**57.8%**	3.1%	10.1%	5.0%	8.8%	1.3%	14.0%	3.36
date	**53.6%**	5.0%	1.9%	4.7%	7.6%	3.2%	23.9%	3.20
beauty	**46.0%**	7.1%	8.3%	3.2%	5.8%	3.1%	26.4%	2.92
janitor	**61.6%**	7.0%	5.7%	6.6%	12.3%	3.4%	3.4%	4.34
cancer	**76.1%**	1.3%	2.6%	5.0%	11.7%	1.0%	2.3%	4.52
ugly	**65.7%**	1.1%	4.6%	15.9%	6.8%	0.8%	5.0%	3.94
Disfigured	**41.1%**	1.6%	7.6%	4.5%	6.6%	0.8%	37.7%	2.52
Brother	**55.6%**	1.3%	3.8%	5.7%	3.3%	6.7%	23.7%	2.92
Disable	**74.3%**	4.5%	4.2%	4.0%	9.9%	1.8%	1.3%	4.52
Dinner	**61.3%**	1.3%	5.5%	4.2%	7.6%	3.6%	16.5%	3.40
Painting	**57.8%**	1.9%	12.9%	5.3%	5.3%	1.8%	14.9%	3.34
obese	**46.0%**	9.9%	6.0%	3.9%	11.7%	0.8%	21.8%	3.27
attractive	**34.3%**	1.5%	2.1%	2.1%	2.9%	3.4%	53.7%	2.04
weight	**64.2%**	1.6%	7.4%	4.0%	6.0%	1.5%	15.2%	3.21
dove	**80.6%**	1.0%	0.6%	1.6%	13.6%	0.8%	1.8%	4.55
cat	**75.4%**	1.0%	2.4%	2.1%	4.4%	0.6%	14.1%	3.45
cow	**84.4%**	0.3%	1.0%	1.9%	7.8%	1.0%	3.6%	3.98
horse	**83.3%**	0.5%	1.9%	1.6%	9.1%	0.6%	2.9%	4.26
cats	**85.3%**	0.5%	1.6%	1.6%	8.3%	0.8%	1.9%	4.26
lion	**84.0%**	0.6%	4.2%	0.6%	8.1%	0.6%	1.8%	4.42
dog	**61.2%**	2.3%	2.1%	3.7%	9.7%	2.8%	18.3%	3.35
doves	**82.3%**	0.6%	1.3%	1.9%	12.0%	0.2%	1.6%	4.47
ruler	**72.1%**	1.1%	10.4%	6.5%	3.9%	1.1%	4.9%	4.08
thumbtack	**77.4%**	1.1%	1.0%	4.7%	7.1%	3.4%	5.2%	3.98
belt	**57.8%**	1.1%	8.7%	3.2%	4.0%	1.0%	24.1%	3.06
CARE	***65.7%***	*2.4%*	*4.8%*	*4.3%*	*7.9%*	*2.0%*	*13.0%*	*3.68*
cheat	1.8%	**72.7%**	1.5%	4.2%	5.3%	10.2%	4.4%	3.88
runner	1.6%	**79.4%**	0.5%	3.2%	2.6%	10.4%	2.3%	4.30
apart	2.1%	**72.2%**	5.7%	2.6%	4.2%	6.1%	7.1%	3.46
soccer	3.4%	**65.9%**	2.6%	3.6%	6.5%	13.4%	4.7%	3.78
card	1.8%	**71.2%**	0.8%	2.4%	4.4%	14.4%	5.0%	3.83
professor	14.8%	**63.3%**	8.8%	4.4%	3.7%	2.4%	2.6%	4.40
referee	2.1%	**73.5%**	2.1%	2.3%	2.1%	15.2%	2.8%	4.21
judge	1.0%	**48.6%**	2.3%	3.1%	2.3%	4.4%	38.4%	2.89
hours	1.3%	**59.8%**	2.1%	6.8%	3.4%	18.6%	7.9%	3.67
tax	3.9%	**58.7%**	1.8%	3.7%	11.5%	19.7%	0.6%	4.77
line	1.9%	**71.2%**	1.5%	10.7%	2.1%	4.4%	8.3%	3.54
vacation	1.6%	**43.7%**	1.8%	6.7%	3.9%	17.6%	24.7%	3.09
FAIRNESS	*3.1%*	***65.0%***	*2.6%*	*4.5%*	*4.3%*	*11.4%*	*9.1%*	*3.82*
competitor	2.9%	4.0%	1.3%	3.6%	3.1%	**78.5%**	6.6%	3.98
team	2.1%	1.6%	0.2%	1.9%	1.5%	**50.4%**	42.3%	2.48
general	2.4%	3.6%	0.5%	6.5%	1.5%	**64.5%**	21.1%	3.25
neighbor	4.2%	5.5%	2.4%	4.9%	1.6%	**55.2%**	26.2%	3.00
Ambassador	7.3%	3.4%	1.8%	13.9%	5.8%	**62.9%**	4.9%	4.16

Items	Care	Fairness	Liberty	Authority	Purity	Loyalty	Not Wrong	Mean Wrongness
Competence	2.9%	2.9%	2.1%	3.9%	0.7%	**63.6%**	23.9%	2.84
Math	4.7%	4.0%	2.4%	3.7%	1.6%	**72.7%**	10.8%	3.45
Cheerleader	10.0%	2.6%	3.6%	5.2%	3.6%	**67.4%**	7.6%	3.62
Faculty	3.1%	3.6%	1.9%	5.0%	2.3%	**64.8%**	19.3%	3.07
Hollywood	3.1%	6.2%	2.9%	7.3%	3.2%	**55.4%**	21.9%	3.39
President	1.5%	1.0%	2.1%	7.3%	1.6%	**40.4%**	46.2%	2.42
Vote	5.5%	2.8%	2.3%	0.8%	2.8%	**47.8%**	38.0%	2.59
Foreigner	7.0%	3.9%	3.4%	10.5%	3.9%	**61.9%**	9.4%	3.91
Wife	1.9%	6.8%	0.6%	1.0%	2.1%	**48.2%**	39.3%	2.49
citizens	1.1%	1.9%	1.5%	5.2%	2.3%	**40.4%**	47.6%	2.40
cheering	9.1%	3.1%	1.5%	3.1%	2.3%	**45.3%**	35.8%	2.65
LOYALTY	*4.3%*	*3.6%*	*1.9%*	*5.2%*	*2.5%*	***57.5%***	*25.1%*	*3.11*
curfew	2.3%	3.9%	0.8%	**74.4%**	2.3%	9.1%	7.3%	3.56
class	2.9%	1.3%	3.1%	**69.3%**	1.8%	3.9%	17.8%	3.22
intern	1.6%	2.9%	6.3%	**62.1%**	2.4%	10.0%	14.6%	2.97
boss	3.7%	1.8%	2.1%	**72.4%**	3.2%	7.0%	9.7%	3.31
mayor	2.4%	3.4%	9.2%	**65.2%**	2.8%	7.1%	9.9%	3.33
teenage	1.3%	4.4%	1.8%	**75.2%**	1.3%	8.7%	7.3%	3.34
employee	5.7%	3.2%	3.7%	**51.5%**	3.2%	25.8%	6.8%	3.61
monitor	7.9%	1.9%	3.6%	**75.5%**	2.1%	5.5%	3.4%	3.75
sermon	2.6%	0.6%	1.6%	**71.4%**	5.5%	7.6%	10.6%	3.39
bench	1.6%	3.1%	1.6%	**69.0%**	0.5%	14.7%	9.4%	3.34
coach	7.1%	1.3%	3.6%	**62.9%**	1.3%	9.9%	13.9%	3.18
sports	1.1%	0.6%	1.3%	**59.6%**	6.6%	9.7%	20.9%	3.12
tv	4.5%	1.8%	2.4%	**54.9%**	2.1%	2.3%	31.9%	2.64
idiot	17.9%	0.5%	2.8%	**65.7%**	5.2%	3.4%	4.5%	3.88
AUTHORITY	*4.5%*	*2.2%*	*3.1%*	***66.4%***	*2.9%*	*8.9%*	*12.0%*	*3.33*
chicken	3.9%	0.8%	1.0%	3.7%	**85.2%**	1.6%	3.7%	4.59
oral	3.7%	1.0%	4.9%	4.7%	**75.4%**	0.8%	9.4%	3.93
philia	8.6%	1.0%	1.6%	3.2%	**77.0%**	1.0%	7.6%	3.99
dece	5.7%	0.2%	1.0%	4.1%	**84.0%**	1.1%	4.1%	4.45
gay	4.4%	1.3%	5.7%	3.4%	**62.7%**	1.0%	21.6%	3.45
morgue	5.2%	0.6%	1.0%	10.6%	**54.1%**	1.1%	27.4%	3.06
cannibal	7.8%	1.0%	1.1%	3.4%	**43.5%**	0.6%	42.5%	2.67
underwear	2.3%	0.6%	1.3%	2.4%	**68.0%**	1.1%	24.2%	3.30
cousins	1.6%	0.6%	1.9%	13.9%	**21.7%**	5.0%	55.1%	2.20
sexdoll	3.6%	0.5%	1.9%	3.1%	**35.5%**	0.8%	54.6%	2.19
PURITY	*4.7%*	*0.8%*	*2.1%*	*5.3%*	***60.7%***	*1.4%*	*25.0%*	*3.38*
switch	1.8%	5.3%	**78.6%**	3.4%	1.3%	4.2%	5.3%	3.83
pilot	4.4%	2.4%	**78.3%**	2.8%	1.1%	1.5%	9.5%	3.42
religion	3.2%	3.6%	**79.7%**	6.8%	1.3%	2.6%	2.8%	4.09
friends	3.6%	4.9%	**78.0%**	2.9%	1.0%	0.6%	9.1%	3.50
clothes	6.8%	4.1%	**75.7%**	2.6%	3.7%	0.5%	6.6%	3.83
pressure	2.4%	12.4%	**74.8%**	2.4%	2.6%	2.9%	2.4%	3.89
daughter	4.9%	1.8%	**77.9%**	1.3%	2.0%	0.8%	11.4%	3.34
sweatsh	2.8%	5.0%	**63.5%**	3.1%	1.8%	1.8%	22.0%	3.14
medicine	3.9%	2.6%	**73.9%**	3.1%	1.1%	1.0%	14.4%	3.16

Items	Care	Fairness	Liberty	Authority	Purity	Loyalty	Not Wrong	Mean Wrongness
business	2.8%	3.4%	**55.6%**	2.8%	0.3%	1.5%	33.7%	2.51
bright	2.1%	4.5%	**74.2%**	5.0%	2.9%	1.5%	9.7%	3.68
LIBERTY	*3.5%*	*4.5%*	***73.7%***	*3.3%*	*1.7%*	*1.7%*	*11.6%*	*3.49*
dirtyca	0.8%	0.8%	0.8%	1.8%	4.1%	0.6%	**91.1%**	1.38
blackwh	1.1%	0.6%	0.5%	0.8%	0.8%	0.3%	**95.8%**	1.16
novel	0.6%	5.7%	0.5%	1.6%	0.3%	1.6%	**89.6%**	1.35
coat	4.5%	0.5%	0.5%	0.8%	1.8%	0.5%	**91.4%**	1.26
hello	4.7%	1.0%	1.8%	4.0%	1.6%	1.1%	**85.8%**	1.43
Neutral	*2.4%*	*1.7%*	*0.8%*	*1.8%*	*1.7%*	*0.8%*	***90.7%***	*1.32*

Note: Items are referred by a keyword in the vignette. The full text of the item can be accessed in the OSF repository. Items in *italics* do not meet the inclusion criteria

and 25% Argentina) or a high school degree (17% Colombia, 19% Peru and 37% Argentina).

3.2.2. Materials and Procedure

Moral Foundation Vignettes. Participants were given the version of the MFV described in Study 1 with versions adapted to their countries. Colombian participants only completed the moral judgement rating questions but not the classification task. Only one neutral item was included in this version of the MFV.

Moral Foundation Questionnaire. The translation of the MFQ 30 into Spanish (Gudino & Fernández-Cárdenas, 2015) was used for all participants. No further adjustment was deemed necessary, since the level of abstraction inherent to this test made the language easily understood by any native Spanish speaker. The MFQ consists of two dimensions, relevance (When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? [*Cuando usted decide que algo está bien o mal, ¿qué tan relevantes son las siguientes consideraciones en su desición?*]) and moral judgment (Please read the following sentences and indicate your agreement or disagreement [Por favor, lea las siguientes afirmaciones e indique qué tan de acuerdo está con ellas]. Both questions require responses from 1 (not at all relevant/strongly disagree [sin relación alguna/totalmente en desacuerdo]) to 6 (extremely relevant/strongly agree [extremadamente relacionado/muy de acuerdo]) and both subscales require evaluating three statements corresponding to five foundations (Care, Fairness, Authority, Loyalty and Purity) for a total of 30 items.

Morality as Cooperation Questionnaire. This instrument was recently developed to measure morality in the framework of the Morality as Cooperation theory (Curry et al., 2019, 2021). The theory and instrument posit that moral judgement and behavior consist of adaptations to distinct cooperation problems that arise during human evolutionary history and can be characterized as distinct game-theoretic scenarios. These cooperation areas (identified with the labels of Family, Heroism, Group, Reciprocity, Deference, Fairness, Property) define seven first-order moral domains that form the basic concerns of moral reasoning. In turn, the MAC-Q implies a seven -factor model of morality that partially overlaps with the structure of the MFQ and the MFV (Curry et al., 2021). It also has two subscales, relevance and judgement and three items per element for a total 21 items per dimension and 42 overall. The relevance question asks participants to rate "to what extent are the following considerations relevant to your thinking? (0–100; not at all relevant, not very relevant, slightly relevant, somewhat relevant, very relevant, extremely relevant)" and the judgement items ask participants to rate "To what extent do you agree with the following statements? (0–100; strongly disagree, disagree, neither agree or disagree, agree, strongly agree)". Since this scale has not been validated in Spanish, we followed a simplified procedure of translation, back translation and cultural adjustment. The final version can be found in the supplemental materials.

The sociodemographic questions were the same as in Study 1. The survey was administered through Qualtrics. Instrument and item order, within instrument, was fully randomized. Participants always saw the sociodemographic questions last.

3.2.3. Results

Overall results are summarized in Table 3. The average successful classification rate was 64% for Peru and 63% for Argentina. Within-foundation classification rates range from 38% to 84% in Peru and from 26% to 88% in Argentina, exhibiting a similar pattern to Colombian responses. Loyalty violations had the lowest perceived wrongness (Peru $M = 3.38$, SD = .46; Argentina $M = 2.99$, SD = .47) and also the lowest classification success rate (Peru 60%; Argentina 54%) in contrast with liberty violations, that had the highest average successful classification rates for both countries (Peru 68%; Argentina 77%). Fairness violations had the highest average wrongness in Peru (M = 3.75, SD = .43) compared with Care infractions in Argentina (M = 3.66, SD = .61). The judged wrongness and classification rates pattern are a mirror image of the Study 1 results. The number of items that failed to meet the inclusion criteria was lower for Peru (26) and higher in Argentina (33). Proportionally, most of these vignettes came from the Loyalty foundation (Peru 7, Argentina 11) and, unlike Colombia, from the Care foundation (Peru 9, Argentina 7).

Table 2. Goodness-of-fit indices of structural models underlying MFV (n = 651; Study 1).

		Ideol.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Peru																		
1	MFV harm	0.05																	
2	MFV justice	0.06	0.72***																
3	MFV loyalty	0.02	0.65***	0.70***															
4	MFV authority	0.09	0.66***	0.74***	0.77***														
5	MFV sanctity	0.08	0.65***	0.66***	0.66***	0.67***													
6	MFV liberty	0.02	0.72***	0.69***	0.56***	0.59***	0.53***												
7	MFQ care	0.01	0.49***	0.42***	0.39***	0.33***	0.36***	0.37***											
8	MFQ fairness	-0.06	0.43***	0.40***	0.34***	0.28***	0.33***	0.34***	0.76***										
9	MFQ loyalty	0.02	0.28***	0.39***	0.50***	0.40***	0.31***	0.25***	0.54***	0.56***									
10	MFQ authority	0.05	0.27***	0.42***	0.47***	0.49***	0.36***	0.21***	0.47***	0.48***	0.69***								
11	MFQ purity	0.10	0.34***	0.41***	0.50***	0.53***	0.49***	0.23***	0.55***	0.50***	0.62***	0.70***							
12	MAC-Q family	0.05	0.19***	0.26***	0.37***	0.35***	0.23***	0.11*	0.33***	0.29***	0.54***	0.47***	0.49***						
13	MAC-Q group	-0.00	0.23***	0.27***	0.30***	0.29***	0.17***	0.15**	0.47***	0.42***	0.50***	0.44***	0.46***	0.64***					
14	MAC-Q reciprocity	0.01	0.27***	0.29***	0.29***	0.30***	0.18***	0.18***	0.50***	0.41***	0.39***	0.40***	0.45***	0.61***	0.73***				
15	MAC-Q heroism	0.05	0.25***	0.32***	0.40***	0.39***	0.26***	0.19***	0.44***	0.36***	0.54***	0.50***	0.54***	0.69***	0.75***	0.73***			
16	MAC-Q deference	0.06	0.21***	0.28***	0.34***	0.37***	0.20***	0.11*	0.35***	0.28***	0.49***	0.56***	0.55***	0.69***	0.68***	0.67***	0.77***		
17	MAC-Q fairness	-0.00	0.30***	0.28***	0.24***	0.24***	0.17***	0.23***	0.45***	0.40***	0.34***	0.32***	0.35***	0.41***	0.60***	0.68***	0.61***	0.60***	
18	MAC-Q property	0.06	0.15**	0.20***	0.09	0.12*	0.13**	0.05	0.32***	0.25***	0.08	0.12*	0.18***	-0.02	0.15**	0.30***	0.15**	0.08	0.42***
	Argentina																		
1	MFV harm	0.06																	
2	MFV justice	0.03	0.61***																
3	MFV loyalty	0.02	0.56***	0.56***															
4	MFV authority	0.05	0.58***	0.66***	0.64***														
5	MFV sanctity	0.12	0.49***	0.42***	0.49***	0.50***													
6	MFV liberty	-0.01	0.69***	0.56***	0.39***	0.41***	0.29***												
7	MFQ care	0.05	0.47***	0.32***	0.27***	0.27***	0.33***	0.34***											
8	MFQ fairness	0.12	0.45***	0.38***	0.27***	0.23***	0.27***	0.39***	0.69***										
9	MFQ loyalty	-0.06	0.30***	0.31***	0.44***	0.41***	0.35***	0.16**	0.37***	0.45***									
10	MFQ authority	0.06	0.22***	0.29***	0.44***	0.52***	0.43***	0.03	0.30***	0.29***	0.56***								
11	MFQ purity	-0.05	0.18***	0.22***	0.38***	0.41***	0.56***	-0.02	0.31***	0.25***	0.50***	0.67***							
12	MAC-Q family	-0.03	0.18***	0.15**	0.28***	0.29***	0.32***	0.00	0.24***	0.18***	0.51***	0.42***	0.48***						
13	MAC-Q group	-0.05	0.23***	0.24***	0.28***	0.25***	0.19***	0.14**	0.39***	0.34***	0.51***	0.24***	0.33***	0.56***					
14	MAC-Q reciprocity	0.08	0.26***	0.29***	0.20***	0.24***	0.21***	0.17***	0.39***	0.39***	0.34***	0.23***	0.27***	0.56***	0.59***				
15	MAC-Q heroism	0.02	0.22***	0.24***	0.30***	0.35***	0.29***	0.07	0.25***	0.21***	0.48***	0.43***	0.44***	0.64***	0.60***	0.62***			
16	MAC-Q deference	0.08	0.15**	0.26***	0.28***	0.41***	0.27***	-0.00	0.17***	0.14**	0.48***	0.57***	0.49***	0.58***	0.51***	0.49***	0.67***		
17	MAC-Q fairness	0.07	0.28***	0.28***	0.16**	0.18***	0.12*	0.23***	0.40***	0.42***	0.26***	0.17***	0.09	0.31***	0.49***	0.56***	0.39***	0.41***	
18	MAC-Q property	0.04	0.23***	0.21***	0.06	0.13**	0.14**	0.19***	0.29***	0.27***	0.09	0.12*	0.08	0.12*	0.15**	0.37***	0.15**	0.19***	0.46***
	Colombia																		
1	MFV harm	0.05																	
2	MFV justice	0.08*	0.74***																

		Ideol.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
3	MFV loyalty	0.14***	0.62***	0.62***															
4	MFV authority	0.15***	0.69***	0.70***	0.72***														
5	MFV sanctity	0.15***	0.57***	0.49***	0.55***	0.58***													
6	MFV liberty	-0.01	0.71***	0.65***	0.47***	0.54***	0.34***												
7	MFQ care	-0.05	0.45***	0.39***	0.31***	0.37***	0.35***	0.32***											
8	MFQ fairness	-0.10	0.36***	0.35***	0.26***	0.27***	0.26***	0.26***	0.74***										
9	MFQ loyalty	0.18**	0.19**	0.30***	0.37***	0.34***	0.29***	0.09	0.46***	0.50***									
10	MFQ authority	0.15*	0.10	0.19**	0.30***	0.30***	0.27***	-0.00	0.41***	0.41***	0.71***								
11	MFQ purity	0.21**	0.16*	0.24***	0.32***	0.38***	0.41***	0.01	0.49***	0.48***	0.62***	0.69***							
12	MAC-Q family	0.06	0.17**	0.24***	0.25***	0.29***	0.32***	-0.01	0.32***	0.27***	0.54***	0.50***	0.48***						
13	MAC-Q group	0.10	0.17*	0.26***	0.16*	0.26***	0.23***	0.04	0.39***	0.36***	0.47***	0.43***	0.40***	0.64***					
14	MAC-Q reciprocity	0.02	0.26***	0.31***	0.18**	0.26***	0.19**	0.07	0.43***	0.44***	0.33***	0.29***	0.30***	0.52***	0.67***				
15	MAC-Q heroism	0.15*	0.15*	0.15*	0.25***	0.26***	0.27***	0.00	0.26***	0.27***	0.46***	0.44***	0.43***	0.66***	0.57***	0.51***			
16	MAC-Q deference	0.16*	0.11	0.21**	0.26***	0.33***	0.25***	0.01	0.21**	0.19**	0.49***	0.52***	0.41***	0.68***	0.56***	0.47***	0.73***		
17	MAC-Q fairness	-0.01	0.32***	0.30***	0.19**	0.26***	0.25***	0.14*	0.39***	0.45***	0.27***	0.27***	0.26***	0.40***	0.50***	0.63***	0.49***	0.51***	
18	MAC-Q property	0.13	0.15*	0.26***	0.15*	0.09	0.15*	0.02	0.37***	0.30***	0.10	0.12	0.15*	0.08	0.19**	0.36***	0.11	0.06	0.43***

Note: CFI: Comparative Fit Index ; TLI: Tucker Lewis Index; IC : Akaike information criterion; BIC: Bayes information criterion; RMSEA: root mean square error of approximation

Table 3. Classification rate and mean wrongness ratings by foundation for Study 2

	PERU								ARGENTINA							
Item	Care	Justice	Liberty	authority	Purity	Loyalty	Not Wrong	Mean Wrongness	Care	Justice	Liberty	Authority	Purity	Loyalty	Not Wrong	Mean Wrongness
CARE	**65%**	**3%**	**5%**	**5%**	**7%**	**3%**	**12%**	3.66	**71%**	2%	3%	2%	8%	2%	13%	3.66
FAIRNESS	4%	**62%**	3%	7%	4%	13%	8%	3.75	3%	**63%**	2%	3%	4%	13%	12%	3.52
LOYALTY	6%	4%	3%	7%	3%	**60%**	18%	3.38	5%	3%	2%	5%	2%	**54%**	28%	2.99
AUTHORITY	5%	3%	3%	**66%**	3%	10%	11%	3.46	5%	3%	2%	**60%**	2%	10%	19%	3.04
PURITY	5%	2%	3%	4%	**64%**	2%	20%	3.59	4%	1%	3%	3%	**58%**	2%	30%	3.26
LIBERTY	5%	6%	**68%**	4%	2%	4%	12%	3.46	4%	3%	**77%**	2%	1%	2%	10%	3.56

We ran exploratory factor analysis per country following the same logic of Study 1. Correlations between items were generally good (KMO Peru = .92, Argentina = .93) and parallel analyses suggested between 5 and 7 factors. Given this and the findings from Study 1, we fitted three factor analyses per country with a Promax rotation. For both countries, the seven-factor solution performed better with adequate fit indices (Peru *RMSEA* = 0.034 [0.031, 0.035], *TLI* = .885; Argentina *RMSEA* = 0.033 [0.031, 0.035], *TLI* = .88).

The composition of the factors however does not match predictions of the MFT and offers a more disordered pictured compared with the Study 1 data. For Peru, the most clearly defined factors are factor 4 (Liberty) and 6 (Care). The rest of the factors exhibit a mix of items from different foundations: Loyalty in factors 5 and 7, authority and fairness items in factor 1, Care and Purity in Factor 2, and Care, Fairness and Purity in factor 3. For Argentina we found that items related to Liberty, Loyalty, Purity and Care loaded onto their own factor neatly (factors 2,3,6 and 7 respectively) while vignettes of Authority and Fairness were mixed in Factor 1 and Care items loaded onto factors 3 and 4 without a clear logic. The most consistent result from the exploratory factor analysis in both Study 1 and 2 is that Liberty items tend to cluster clearly together as well as Care items. Items that loaded weakly on any factor (<.3) matched the items that did not meet the classification thresholds, so no other items were flagged at this stage.

We conducted confirmatory factor analysis with six and seven factors with the hypothesized structure by country excluding items identified in the prior analysis. Summary of the fit indices for these models can be seen in Table 4. For both countries, a likelihood ratio test revealed no significant differences in fit between the six and seven factor solution, even though the seven-factor model performs slightly better in most indicators. We decided to retain the seven-factor solution, in line with Study 1 results and Marques et al (2020).

The MFQ and the MAC-Q were also analyzed with a series of exploratory and confirmatory factor analyses (see supplemental materials). Solutions for both the MFQ consistently point out to a two-factor solution that roughly coincided with distinction between individualizing and binding foundations. The MAC-Q differs between countries ranging from four to seven factor solutions. We also estimated scores for the MFV, MFQ and MAQ-Q by averaging ratings by participant and by foundation/moral element. For the MFV, this was restricted only to the items that were retained in the previous analyses per country. We then calculated correlations between all subscales scores and also included our measure of ideology. Results are presented in Table 5.

We observed strong correlations between foundation scores across the MFV and the MFQ for all countries, as expected (e.g MFV Care and MFQ Care in Peru: $r(397) = .49$, $p < 0.001$). Scores of the MAC-Q are also positively correlated with MFV scales scores. For example, Authority ratings correlate significantly with family (between .29 and .35), deference (between .37 and .43) and group (between .26 and .29), elements of the MAC that overlap with the Authority foundation. Similarly, Loyalty MFV scores are significantly correlated with the MAC-Q group scores (between .16 and .37), which share aspects in their definition. Correlation between conservatism and MFV scores were also consistent with the development of the original instrument only for Colombia (See Table 5). Surprisingly, conservatism ratings are not correlated with almost any of the scales scores.

Final reliability estimates for the MFV per Foundation for both countries are excellent (Peru: Care (α =.92, ω = .93), Fairness (α =.85, ω =.88), Authority (α =.90, ω =.91), Loyalty (α = .84, ω = .86), Purity (α = .80, ω = .85) and Liberty (α = .85, ω = .88); Argentina: Care (α =.92, ω = .93), Fairness (α =.80, ω =.83), Authority (α =.80, ω =.86), Loyalty (α = .74, ω = .78), Purity (α = .70, ω = .74) and Liberty (α = .87, ω = .89)).

3.2.4. Discussion

Study 2 provided further evidence that the seven-factor structure is adequate to capture moral judgement in two additional countries. The validation process resulted in versions of the MFV that are adapted for three countries that, however, implied retaining a smaller number of vignettes (Peru 64; Argentina 58, Colombia 60) compared to the original instrument. We again found that participants failed to recognize the intended foundation in a large proportion of items, particularly for Loyalty vignettes. . The final set of items is available on the project's OSF page, where it is also evident that many of the excluded items are consistent across countries. Alternative measures of moral judgment correlate well with the adapted versions of the MFV although, surprisingly, reported political ideology is not associated with any of these measures

Phase 3. Cross-country Comparisons

According to a recent metric of cultural distance (Muthukrishna et al., 2020), South American countries and the US are relatively close (Peru = .090 [.087, .094], Colombia = .101, [.0987, .106], Argentina = 0.71., [069, .075]) compared to rest of world though they are considered non-WEIRD societies (Henrich et al., 2010) and also part of the Global South. Adaptation of instruments like the MFV are essential for improving our understanding of cross-cultural differences and the way psychometric instruments can be used to measure for moral regularity and diversity (Apicella et al., 2020). With this mind, we conducted several cross-country comparisons to assess the cross-cultural validity of the MFV.

3.3. Study 3. Measurement Invariance and Comparisons Across Countries

Invariance was tested using the *lavaan* R package (Rosseel, 2012). Two sets of analyses were performed, one with the six-factor structure of the original instrument and another one with the seven-factor structure identified in Studies 1 and 2. We used data from Studies 1 and 2 and also the original US data but restricted the analysis to the subset of items common to all versions (52).

Table 4. Goodness-of-fit indices of structural models underlying MFV (n = 1032; Study 2)

Country	Model	χ^2	df	CFI	TLI	AIC	BIC	RMSEA	90% CI
Peru	6 factors	4372.46	2000	0.81	0.80	64896.3	65474.7	0.052	[0.052 0.057]
	7 Factors (Emotional and physical care)	4168.09	1994	0.82	0.82	64703.9	65306.2	0.046	[0.052 0.055]
Argentina	6 factors	3774.541	1580	0.76	0.76	62774.8	63297.7	0.059	[0.057 0.061]
	7 Factors (Emotional and physical care)	3375.515	1574	0.80	0.80	62387.7	62934.6	0.053	[0.051 0.056]

Note: CFI: Comparative Fit Index; TLI: Tucker Lewis Index; IC: Akaike information criterion; BIC: Bayes information criterion; RMSEA: root mean square error of approximation

Table 5. Correlations between subscales of the conservatism, MFV, MFQ and MAC-Q

		Ideol.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Peru																
1	MFV harm	0.05															
2	MFV loyalty	0.02	0.65***														
3	MFV authority	0.09	0.66***	0.77***													
4	MFV Purity	0.08	0.65***	0.66***	0.67***												
5	MFV liberty	0.02	0.72***	0.56***	0.59***	0.53***											
6	MFQ care	0.01	0.49***	0.39***	0.33***	0.36***	0.37***										
7	MFQ fairness	-0.06	0.43***	0.34***	0.28***	0.33***	0.34***	0.76***									
8	MFQ loyalty	0.02	0.28***	0.50***	0.40***	0.31***	0.25***	0.54***	0.56***								
9	MFQ authority	0.05	0.27***	0.47***	0.49***	0.36***	0.21***	0.47***	0.48***	0.69***							
10	MFQ purity	0.1	0.34***	0.50***	0.53***	0.49***	0.23***	0.55***	0.50***	0.62***	0.70***						
11	MAC-Q family	0.05	0.19***	0.37***	0.35***	0.23***	0.11*	0.33***	0.29***	0.54***	0.47***	0.49***					
12	MAC-Q group	0	0.23***	0.30***	0.29***	0.17***	0.15**	0.47***	0.42***	0.50***	0.44***	0.46***	0.64***				
13	MAC-Q heroism	0.05	0.25***	0.40***	0.39***	0.26***	0.19***	0.44***	0.36***	0.54***	0.50***	0.54***	0.69***	0.75***			
14	MAC-Q deference	0.06	0.21***	0.34***	0.37***	0.20***	0.11*	0.35***	0.28***	0.49***	0.56***	0.55***	0.69***	0.68***	0.77***		
15	MAC-Q fairness	0	0.30***	0.24***	0.24***	0.17***	0.23***	0.45***	0.40***	0.34***	0.32***	0.35***	0.41***	0.60***	0.61***	0.60***	
16	MAC-Q property	0.06	0.15**	0.09	0.12*	0.13**	0.05	0.32***	0.25***	0.08	0.12*	0.18***	-0.02	0.15**	0.15**	0.08	0.42***
	Argentina																
1	MFV harm	0.06															
2	MFV loyalty	0.02	0.56***														
3	MFV authority	0.05	0.58***	0.64***													
4	MFV purity	0.12	0.49***	0.49***	0.50***												
5	MFV liberty	-0.01	0.69***	0.39***	0.41***	0.29***											
6	MFQ care	0.05	0.47***	0.27***	0.27***	0.33***	0.34***										
7	MFQ fairness	0.12	0.45***	0.27***	0.23***	0.27***	0.39***	0.69***									
8	MFQ loyalty	-0.06	0.30***	0.44***	0.41***	0.35***	0.16**	0.37***	0.45***								
9	MFQ authority	0.06	0.22***	0.44***	0.52***	0.43***	0.03	0.30***	0.29***	0.56***							
10	MFQ purity	-0.05	0.18***	0.38***	0.41***	0.56***	-0.02	0.31***	0.25***	0.50***	0.67***						

		Ideol.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	MAC-Q family	-0.03	0.18***	0.28***	0.29***	0.32***	0	0.24***	0.18***	0.51***	0.42***	0.48***					
12	MAC-Q group	-0.05	0.23***	0.28***	0.25***	0.19***	0.14**	0.39***	0.34***	0.51***	0.24***	0.33***	0.56***				
13	MAC-Q heroism	0.02	0.22***	0.30***	0.35***	0.29***	0.07	0.25***	0.21***	0.48***	0.43***	0.44***	0.64***	0.60***			
14	MAC-Q deference	0.08	0.15**	0.28***	0.41***	0.27***	0	0.17***	0.14**	0.48***	0.57***	0.49***	0.58***	0.51***	0.67***		
15	MAC-Q fairness	0.07	0.28***	0.16**	0.18***	0.12*	0.23***	0.40***	0.42***	0.26***	0.17***	0.09	0.31***	0.49***	0.39***	0.41***	
16	MAC-Q property	0.04	0.23***	0.06	0.13**	0.14**	0.19***	0.29***	0.27***	0.09	0.12*	0.08	0.12*	0.15**	0.15**	0.19***	0.46***
	Colombia																
1	MFV harm	0.05															
2	MFV loyalty	0.14***	0.62***														
3	MFV authority	0.15***	0.69***	0.72***						X|							
4	MFV purity	0.15***	0.56***	0.55***	0.57***												
5	MFV liberty	-0.01	0.71***	0.47***	0.54***	0.34***											
6	MFQ care	-0.05	0.45***	0.31***	0.37***	0.34***	0.32***										
7	MFQ fairness	-0.1	0.36***	0.26***	0.27***	0.24***	0.26***	0.74***									
8	MFQ loyalty	0.18**	0.19**	0.37***	0.34***	0.28***	0.09	0.46***	0.50***								
9	MFQ authority	0.15*	0.1	0.30***	0.30***	0.26***	0	0.41***	0.41***	0.71***							
10	MFQ purity	0.21**	0.16*	0.32***	0.38***	0.41***	0.01	0.49***	0.48***	0.62***	0.69***						
11	MAC-Q family	0.06	0.17**	0.25***	0.29***	0.30***	-0.01	0.32***	0.27***	0.54***	0.50***	0.48***					
12	MAC-Q group	0.1	0.17*	0.16*	0.26***	0.21**	0.04	0.39***	0.36***	0.47***	0.43***	0.40***	0.64***				
13	MAC-Q heroism	0.15*	0.15*	0.25***	0.26***	0.25***	0	0.26***	0.27***	0.46***	0.44***	0.43***	0.66***	0.57***			
14	MAC-Q deference	0.16*	0.11	0.26***	0.33***	0.24***	0.01	0.21**	0.19**	0.49***	0.52***	0.41***	0.68***	0.56***	0.73***		
15	MAC-Q fairness	-0.01	0.32***	0.19**	0.26***	0.23***	0.14*	0.39***	0.45***	0.27***	0.27***	0.26***	0.40***	0.50***	0.49***	0.51***	
16	MAC-Q property	0.13	0.15*	0.15*	0.09	0.15*	0.02	0.37***	0.30***	0.1	0.12	0.15*	0.08	0.19**	0.11	0.06	0.43***

Note: * < 0.1, **<0.05, ***<0.01

Table 6. Measurement invariance modeling

	Model	χ^2	(df)	CFI	RMSEA	ΔCFI	ΔRMSEA	SRMR	AIC
6 Factors	Configural	16059.5	(6320)	0.817	0.055	--	--	0.060	312016.39
	Metric	16327.6	(6476)*	0.815	0.054	0.01	0.00	0.065	311972.49
	Scalar	19528.1	(6632)*	0.758	0.061	0.06	0.008	0.070	314860.93
7 Factors	Configural	14496.9	(6296)	0.840	0.050	--	--	0.057	310501.78
	Metric	14710.3	(6449)*	0.845	0.050	0.02	0.00	0.061	310409.11
	Scalar	17942.2	(6602)*	0.787	0.058	0.06	0.007	0.067	313335.00

Note. *p < 0.001, df = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square residual; AIC = Akaike Information Criteria

We fitted a set of multigroup structural equation models with a Maximum Likelihood estimator. Since about 2% of the data was missing (between 1 and 3 data points out of the 2066 observations), and descriptive statistics did not suggest any pattern, we assumed that data was missing at random and allowed Lavaan to compute likelihood for missing cases without having to delete them casewise. Latent variables were scaled by setting variances to 1 and factor loadings were estimated freely for all variables. We first tested for configural invariance, to determine whether the factor structure was invariant across countries. We then tested for metric invariance, holding loadings equal across countries. We tested for scalar invariance by additionally holding intercepts constant across countries. We finally performed a likelihood ratio test to compare model performance. Results are presented in Table 6.

The analysis suggests that there is only weak evidence of configural invariance, slightly better for the seven versus the six-factor structure (LR test). The analysis weakly suggests that the overall MFV structure is supported across countries. Metric and scalar models performed worse (as indicated by the significant $\chi 2$ difference tests) suggesting that participants are treating items differently and vignettes are contributing to each Foundation differently across countries (Davidov et al., 2014; Putnick & Bornstein, 2016) and so direct point comparisons are not meaningful.

3.4. Differential Item Functioning of Vignettes

To further explore the results of the measure invariance testing, we conducted a Differential Item Functioning Analysis (DIF) by Foundation (See supplemental materials for details). DIF is a technique used in the context of Item Response Theory that shifts focus from the test structure to explain how the latent trait determines the probability of response. It has been used widely in educational and clinical psychology to identify biases at the item level (Gonzalez & Pelham, 2020; Klieme & Baumert, 2001) and more recently in cultural comparisons outside psychometrics (Davidov et al., 2012).

We conducted DIF using the R package *lordif* (Choi et al., 2016) which enables running ordinal logistic regressions with IRT-based trait scores and can be used to determine uniform (differences between groups are constant for all levels of the variable) and non-uniform (differences are not constant) diferentail item functioning. The analysis revealed uniform DIF in 45 out of the 52 items considered and non-uniform DIF in 12. To understand these differences let's consider the case of Fairness, and one of its items, *tax*[3]. The trait distribution shows that although there is large overlap in the way participants judge Fairness vignettes, Argentinian participants reported lower scores than Colombian participants, which might suggest varying levels of sensibility to violations across foundations (see Figure 1). The true score function for the item "tax" reveals that this vignette is consistently considered wrong at low levels of the trait (e.g. sense of Fairness) for all countries but that Argentinians more readily judge this vignette as morally wrong compared with Peruvians and Americans (See Figure 1B). It also shows a steeper slope for the Argentinian sample, implying that smaller changes in sensitivity to Fairness are associated with large changes to judged wrongness (The item "tax" exhibits uniform and non-uniform DIF).

4. General Discussion

In this paper we validated the Moral Foundation Vignettes for Latin-American Spanish by using a large and diverse sample from three countries. We selected the set of vignettes that adequately capture the structure of the MFT and are also sensitive to moral concerns in the region and by country, for future use. The structure of the instrument is coherent with a seven-factor solution, derived and consistent with the six moral foundation taxonomy that underlies the original development of the instrument. We believe the MFV is a powerful tool to examine every day moral judgement.

Unlike research with the MFQ, we did not find a strong association between political ideology and the MFV scores, except for Colombia. MFQ and MAC-Q scores provided a good measure of content convergence and generally point to a good construct validity. Cross national comparisons suggest that the MFV structure and content varies impor-

3 You see a politician using federal tax dollars to build an extension on his home.

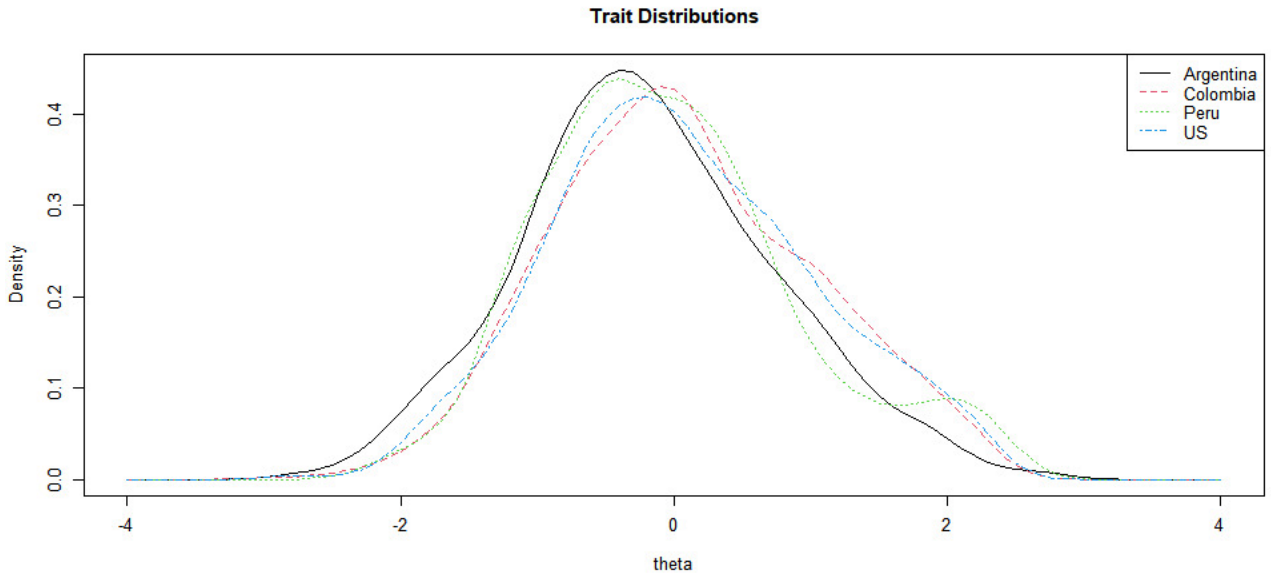

Figure 1A. Trait distribution for Fairness vignettes

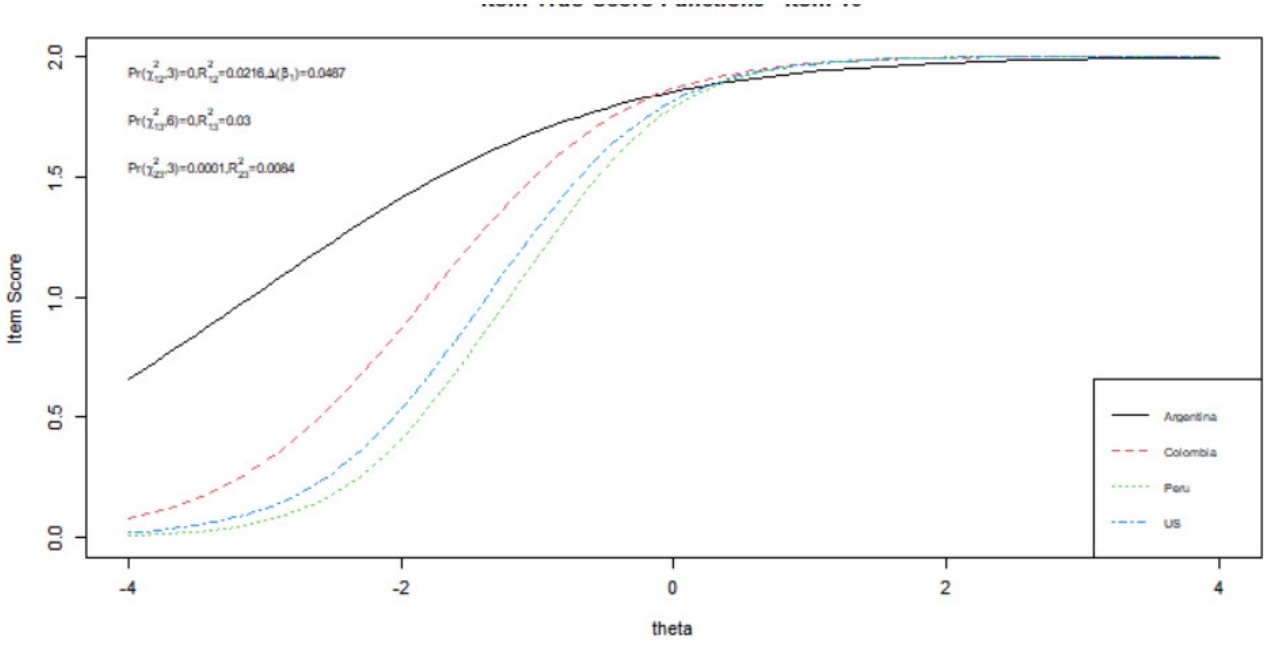

Figure 1B. True score function for item "tax"

tantly by country and that new vignettes should be developed to improve cultural adaptation.

In the MFV participants had to rate the moral wrongness of situations defined according to the MFT. Subsequently though they have to classify the violation using only sparse information about the restricted definition of the and their own cognitive representation of this concept. Since the structural analysis and cross-national comparisons were restricted to these moral judgements, we cannot be sure the validated versions of the MFV adequately represent neither the intended conceptual space of the MFT theory nor lay intuitions about Foundations. By restricting the number of vignettes with face validity, the MFV has run into a variation of the attenuation paradox (Clark & Watson, 2019): reducing the number of items increases the consistency of the items at the cost of restricting content validity.

The case of loyalty is very telling in this regard. Proportionally, the largest number of vignettes discarded came from this foundation, since participants did not consider them either wrong or as violating a loyalty norm. However, the instruction in the classification task was overly restricted (it violates norms of loyalty (e.g. betrayal of a group)) and presumably does not match the cognitive representation of a concept that encompasses more than individual-to-group relations, as the MFT implies. In a related study, we explore this hypothesis (Murray et al., 2024) and

found that constructing items in the loyalty foundation out of culturally specific prototypes increases accuracy scores to similar (or higher) levels as those observed with respect to other foundations (e.g. care, liberty).

In the context of current efforts to refine tools to do cross cultural research (Apicella et al., 2020) we show how to integrate several of the methods available to those interested in this area. Despite our samples being diverse, online samples in Latin-American might not be representative of the countries, given that the use of a polling firm relies on uneven access to internet in the region. This means that political orientation, for example is tilted right or left, because over or underrepresentation of the individuals sampled. In our case, for example, for Colombia and Peru there were more people who reported leaning right.

Contributions

Contributed to conception and design: WJL, GC, SM & SA

Contributed to acquisition of data: WJL, GC, & SA

Contributed to analysis and interpretation of data: WJL & GC

Drafted and/or revised the article: WJL, GC, SM, SA

Approved the submitted version for publication: WJL, GC, SM, SA

Competing Interests

We have no conflict of interests to declare.

Funding Information

Funded by the James S. McDonnell Foundation (https://doi.org/10.37717/2020-1200)

Data Accessibility Statement

All presentation materials and participant data can be found on this paper's project page on the project

https://osf.io/mjtxw/

Submitted: July 21, 2023 PDT. Accepted: December 13, 2024 PDT. Published: April 18, 2025 PDT.

References

Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, *41*(5), 319–329. https://doi.org/10.1016/j.evolhumbehav.2020.07.015

Atari, M., Graham, J., & Dehghani, M. (2020). Foundations of morality in Iran. *Evolution and Human Behavior*, *41*(5), 367–384. https://doi.org/10.1016/j.evolhumbehav.2020.07.014

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, *125*(5), 1157–1188. https://doi.org/10.1037/pspp0000470

Barrett, H. C. (2020). Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution and Human Behavior*, *41*(5), 445–453. https://doi.org/10.1016/j.evolhumbehav.2020.05.006

Beal, B. (2020). What are the irreducible basic elements of morality? A critique of the debate over monism and pluralism in moral psychology. *Perspectives on Psychological Science*, *15*(2), 273–290. https://doi.org/10.1177/1745691619867106

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Chadwick, R. A., Bromgard, G., Bromgard, I., & Trafimow, D. (2006). An index of specific behaviors in the moral domain. *Behavior Research Methods*, *38*(4), 692–697. https://doi.org/10.3758/BF03193902

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2016). *lordif: Logistic Ordinal Regression Differential Item Functioning using IRT*. https://cran.r-project.org/package=lordif

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Clifford, S. (2017). Individual Differences in Group Loyalty Predict Partisan Strength. *Political Behavior*, *39*(3), 531–552. https://doi.org/10.1007/s11109-016-9367-3

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: a standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, *47*(4), 1178–1198. https://doi.org/10.3758/s13428-014-0551-2

Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0190954

Crone, D. L., Rhee, J. J., & Laham, S. M. (2021). Developing brief versions of the Moral Foundations Vignettes using a genetic algorithm-based approach. *Behavior Research Methods*, *53*, 1179–1187. https://doi.org/10.3758/s13428-020-01489-y

Curry, O. S., Alfano, M., Brandt, M. J., & Pelican, C. (2021). Moral Molecules: Morality as a Combinatorial System. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021-00540-x

Curry, O. S., Jones Chesters, M., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Davidov, E., Schmidt, P., Billiet, J., & Meuleman, B. (2012). *Cross-cultural analysis*. Routledge Academic. https://doi.org/10.4324/9780203882924

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N., Vaisey, S., Iliev, R., & Graham, J. (2016). Purity Homophily in Social Networks. *Journal of Experimental Psychology. General*, *145*(2), 366–375. https://doi.org/10.1109/CFIS.2015.7391644

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, *151*(July), 109547. https://doi.org/10.1016/j.paid.2019.109547

Fitouchi, L., André, J. B., & Baumard, N. (2023). Moral disciplining: The cognitive and evolutionary foundations of puritanical morality. *Behavioral and Brain Sciences*, *46*, e293. https://doi.org/10.1017/S0140525X22002047

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Gonzalez, O., & Pelham, W. E. (2020). When Does Differential Item Functioning Matter for Screening? A Method for Empirical Evaluation. *Assessment*, *28*(2), 446–456. https://doi.org/10.1177/1073191120913618

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral foundations theory: On the advantages of moral pluralism over moral monism. *Atlas of Moral Psychology*, *211*, 211–222.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., Ditto, P. H., & Gilligan, C. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Gray, K., & Keeney, J. E. (2015a). Disconfirming moral foundations theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science*, *6*(8), 874–877. https://doi.org/10.1177/1948550615592243

Gray, K., & Keeney, J. E. (2015b). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, *6*(8), 859–868. https://doi.org/10.1177/1948550615592241

Gudino, S., & Fernández-Cárdenas, J. (2015). Propiedades psicométricas del cuestionario de fundamentos morales en alumnos de bachillerato: Un estudio exploratorio. *Enseñanza e Investigación En Psicología*, *20*(2), 130–139.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814. https://doi.org/10.1037/0033-295X.108.4.814

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, *102*(6), 1318. https://doi.org/10.1037/a0026545

Hoover, J., Portillo-Wightman, G., Yeh, L., Havaldar, S., Davani, A. M., Lin, Y., Kennedy, B., Atari, M., Kamel, Z., Mendlen, M., Moreno, G., Park, C., Chang, T. E., Chin, J., Leong, C., Leung, J. Y., Mirinjian, A., & Dehghani, M. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*, *11*(8), 1057–1071. https://doi.org/10.1177/1948550619876629

Hruschka, D. J. (2020). "What we look with" is as important as "What we look at." *Evolution and Human Behavior*, *41*(5), 458–459. https://doi.org/10.1016/j.evolhumbehav.2020.07.011

Iliescu, D. (2017). *Adapting tests in linguistic and cultural situations*. Cambridge University Press. https://doi.org/10.1017/9781316273203

Isler, O., Yilmaz, O., & Doğruyol, B. (2021). Are we at all liberal at heart? High-powered tests find no effect of intuitive thinking on moral foundations. *Journal of Experimental Social Psychology*, *92*(August 2020), 104050. https://doi.org/10.1016/j.jesp.2020.104050

Iurino, K., & Saucier, G. (2020). Testing measurement invariance of the Moral Foundations Questionnaire across 27 countries. *Assessment*, *27*(2), 365–372. https://doi.org/10.1177/1073191118817916

Klieme, E., & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, *16*(3), 385. https://doi.org/10.1007/BF03173189

Marques, L. M., Clifford, S., & Cabeza, R. (2020). Translation and validation of the Moral Foundations Vignettes (MFVs) for the Portuguese language in a Brazilian sample. *Judgement and Decision Making*, *15*(1), 149–158. https://doi.org/10.1017/S1930297500006963

Murray, S., Carmona Díaz, G. M., Vega-Plazas, L. S., Jiménez-Leal, W., & Amaya, S. (2024). Loyalty from a personal point of view: A cross-cultural prototype study of loyalty. *Journal of Experimental Psychology: General*.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science*, *31*(6), 678–701. https://doi.org/10.1177/0956797620916782

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Rosseel, Y. (2012). {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Wagemans, F. M. A., Brandt, M. J., & Zeelenberg, M. (2018). Disgust sensitivity is primarily associated with purity-based moral judgments. *Emotion*, *18*(2), 277–289. https://doi.org/10.1037/emo0000359

Supplementary Materials

Peer Review Communication

Download: https://collabra.scholasticahq.com/article/128178-validation-of-the-moral-foundation-vignettes-in-latin-america-the-scope-of-moral-foundations-through-the-lens-of-an-instrument/attachment/261301.docx?auth_token=7YlM00oC1xP-Da47i-9T