The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures[☆],[☆☆]

Burak Doğruyol[a],[*], Sinan Alper[b], Onurcan Yilmaz[c]

[a] Department of Psychology, Altınbaş University, Istanbul, Turkey
[b] Department of Psychology, Yasar University, Izmir, Turkey
[c] Department of Psychology, Kadir Has University, Istanbul, Turkey

ABSTRACT

Although numerous models attempted to explain the nature of moral judgment, moral foundations theory (MFT) led to a paradigmatic change in this field by proposing pluralist "moralities" (care, fairness, loyalty, authority, sanctity). The five-factor structure of MFT is thought to be universal and rooted in the evolutionary past but the evidence is scarce regarding the stability of this five-factor structure across diverse cultures. We tested this universality argument in a cross-cultural dataset of 30 diverse societies spanning the WEIRD (Western, educated, industrialized, rich, democratic) and non-WEIRD cultures by testing measurement invariance of the short-form of the moral foundations questionnaire. The results supported the original conceptualization that there are at least five diverse moralities although loadings of items differ across WEIRD and non-WEIRD cultures. In other words, the current research shows for the first time that the five-factor structure of MFT is stable in the WEIRD and non-WEIRD cultures.

1. Introduction

There have been theoretical models trying to explain the content and cognitive underpinnings of moral judgment (Curry, Jones Chesters, & Van Lissa, 2019; Haidt, 2001; Kohlberg, 1969; Piaget, 1965; Shweder, Much, Mahapatra, & Park, 1997). One of these approaches is the Moral Foundations Theory (MFT), which argues that our moral understanding is a result of our evolved psychology (Graham et al., 2013; Haidt, 2007). According to this theory, there are at least five basic moral foundations, each distinctly evolved to solve a specific adaptive problem in our evolutionary past. *Care/harm* is defined as the caring behavior toward other group members who are in need of protection. *Fairness/justice* is associated with the sensitivity for inequality and the motivation to maintain justice within the group. *Loyalty/betrayal* is related to protecting the interests of one's own group, favoring one's own group member, and discriminating against out-groups. *Authority/subversion* is the desire to maintain the hierarchical structure in the group, and respect for those who are higher in authority. *Sanctity/degradation* is related to the suppression of carnal desires, a motivation to be pure both physically and spiritually and to avoid infectious diseases. Graham, Haidt, and Nosek (2009) describe the principles of care and

fairness as the individualizing foundations since they are all related to individual rights, while the other three foundations are defined as the binding foundations since they bind people together as groups. While liberals perceive only individualizing foundations as morally relevant, conservatives give relatively equal importance to the five foundations and value binding foundations greater than liberals (Graham et al., 2009).

Although this approach to morality received numerous empirical support (see for a review Graham et al., 2013) and became popular, it has also received several criticisms. First and foremost, the statistical fit values of the moral foundations questionnaire (MFQ), which was designed to measure the theoretical framework of MFT, are below the conventional criteria. Graham et al. (2011) ran confirmatory factor analyses to the English version of MFQ in order to determine whether the 5-factor model of MFT fits data better than alternative models and showed that the 5-factor model fits the data better than the two-factor (individualizing vs. binding) and single factor models. Furthermore, independent standardization studies in different cultures have replicated this initial finding of Graham et al. (2009). However, in all these studies, fit values were below the conventional criteria (e.g., Yalçındağ et al., 2017). Davis et al. (2016) even showed that MFQ does

not work well in a US sample of mostly black participants. A very recent cross-cultural study using the short form of MFQ in 27 different cultures also showed measurement non-invariance across cultures (Iurino & Saucier, 2019). In other words, there is some evidence suggesting that the five-factor model proposed by the theory is not cross-culturally valid. Therefore, we need more work to determine the cross-cultural stability of the model proposed by MFT, and determine boundary conditions regarding cross-cultural differences.

An approach to characterize the cultural context representing recruited in mainstream research was proposed by Henrich, Heine, and Norenzayan (2010). According to this approach, the majority of the samples used in behavioral science studies include participants from Western, educated, industrialized, democratic and rich (WEIRD) countries, but these five characteristics represent a very small minority of the world. Similarly, the vast majority of the data were collected at YourMorals.org through the English version of the MFQ in various parts of the world. However, on this platform, participants participate in studies on a voluntary basis, not on any incentive method, which violates the random selection procedure. In addition to this, although the data is gathered from different countries all around the world, since the English version of the scale was the single option, only English speaking participants recruited to the project, indicating that the participants might be the most WEIRD of their home countries. Therefore, there is a need to test the predictions of the MFT in non-WEIRD cultures with locally translated tools.

Although the main predictions of MFT were replicated in both WEIRD (e.g., Davies, Sibley, & Liu, 2014; Métayer & Pahlavan, 2014; Nilsson & Erlandsson, 2015) and non-WEIRD (Berniūnas, Dranseika, & Sousa, 2016; Yilmaz, Harma, Bahçekapili, & Cesur, 2016; Zhang & Li, 2015) cultures, the question of whether the five-factor model proposed by MFT are stable across WEIRD and non-WEIRD cultures has not been previously tested. In this study, we used the cross-cultural dataset collected for Many Labs 2 project (Klein et al., 2018) and aim to test the measurement invariance of the short form of the MFQ across WEIRD and non-WEIRD cultures in 30 politically different countries. Before measurement invariance tests, separate CFAs will be conducted to test the goodness of fit for two and five-factor structures in each cultural groups (i.e., WEIRD, non-WEIRD). In order to test the measurement invariance, the procedure proposed by Muthén and Muthén (1998-2012) will be used. First, it will be examined whether the five-factor structure proposed by the theory is cross-culturally stable (i.e., configural invariance). Second, whether the item loadings (i.e., metric invariance) are similar will be examined. Finally, we will test whether the item means (i.e., scalar invariance) are cross-culturally similar.

2. Method

2.1. Participants

Data were retrieved from the Many Labs 2 Project (Klein et al., 2018) for the current study. Many Labs 2 Project includes a series of replication studies from 36 countries, consisting of 15,305 participants. Originally, the project includes two slates, however, the Moral Foundations Questionnaire was only used in the first slate. Therefore, for the current purposes, only data from the first slate were analyzed, consisted of 7263 participants from 30 countries. Of the participants, 66.43% were female. The average age of the whole sample was 21.91 ($SD = 3.27$).

2.2. Measures

Moral Foundations Questionnaire (Graham et al., 2011) was used in Many Labs 2 Project Slate 1 including three items for each moral foundations measuring care, fairness, loyalty, authority, and sanctity. Participants responded on a six-point Likert-type scale (1 = not at all relevant; 6 = extremely relevant). Cronbach's alpha scores for the each

Table 1
Descriptive statistics and Cronbach's alphas for moral foundations.

	WEIRD			Non-WEIRD			Total		
	Mean	SD	α	Mean	SD	α	Mean	SD	α
Care	5.01	0.81	0.74	4.63	1.01	0.78	4.90	0.89	0.76
Fairness	4.81	0.80	0.68	4.62	0.96	0.77	4.76	0.85	0.72
Loyalty	4.14	0.95	0.62	4.09	0.97	0.60	4.12	0.96	0.61
Authority	3.62	0.87	0.49	3.75	0.91	0.48	3.66	0.88	0.49
Purity	4.06	1.01	0.71	4.04	1.02	0.72	4.05	1.01	0.71
Individualizing	4.91	0.73	0.81	4.63	0.91	0.86	4.83	0.79	0.83
Binding	3.94	0.76	0.78	3.96	0.79	0.79	3.94	0.76	0.78

foundations were sufficient except for authority foundation (0.76, 0.72, 0.61, 0.49, 0.71 respectively). Besides, care and fairness foundations were averaged to form a composite individualizing foundations while loyalty, authority, and sanctity were combined as binding foundations scores. Both individualizing and binding foundations had satisfactory reliability scores (0.83, and 0.78; respectively). Cronbach's alpha scores and descriptive statistics of each moral foundation for WEIRD, non-WEIRD samples, and whole sample is presented in Table 1.

WEIRDness (i.e., Western, Educated, Industrialized, Rich, and Democratic; Henrich et al., 2010) scores of the countries were calculated by scoring each of the five dimensions (Klein et al., 2018). Then, the combined score was dichotomized by using the average score of WEIRDness scores as a cut-off (see https://osf.io/b7qrt/). Countries below the average were coded as non-WEIRD whereas countries above the average were coded as WEIRD (Table 2). In the current study, we use the same strategy with Many Labs 2 project to score WEIRDness of each country.

2.3. Analysis strategy

All CFA models were conducted in *Mplus* Version 7. The full-information maximum-likelihood method was applied to handle missing data. Before the main analyses, cases were excluded with more than half of the data missing and detected to be multivariate outliers on MFQ items using Mahalanobis distance ($\chi^2(15) = 37.697$, $p = .001$). Overall, 295 (4%) cases were eliminated from the dataset. Main analyses were conducted on 6968 participants ($N_{WEIRD} = 4971$). Correlations among moral foundations tested in CFA models are presented in Table 3.

To test CFA models, raw data was used as input. Normal theory weighted least squares χ2 was used for the evaluation of model fit. Furthermore, following Hu and Bentler's (1999) two-index presentation strategy, comparative fit index (CFI), Tucker-Lewis index (TLI), the standardized root mean square residual (SRMR), and the root-mean-square-error of approximation (RMSEA) were used to evaluate the model fit. Values close to 0.06 for RMSEA, values close to 0.95 for CFI and TLI, and values close to 0.08 for SRMR are indicative of good fit (Hu & Bentler, 1999). The χ2-difference-test (Δχ2) was applied to compare the relative model fit. This test was applied to measurement invariance tests across the two-factor model and five-factor model separately since the test requires nested models in which competing models have the same number of parameters (Cheung & Rensvold, 2002).

Latent variables were scaled by fixing one of the indicator's factor loading at 1. Furthermore, since models with freely estimated latent means are not identified and make it harder to detect the source of noninvariance, latent means were fixed to be zero on one group.

2.4. Statistical models

First, to validate factor structure, a series of Confirmatory Factor Analyses (CFA) was conducted on two and five-factor models of the

Table 2
List of countries included in slate 1 of many labs 2 project (in alphabetical order).

WEIRD countries	Austria, Belgium, Canada, Chile, Czech Republic, France, Germany, Hungary, New Zealand, Poland, Portugal, Spain, Sweden, Switzerland, The Netherlands, UK, USA
Non-WEIRD countries	Brazil, China, Costa Rica, Hong Kong (China), India, Japan, Mexico, Serbia, South Africa, Taiwan (China), Turkey, UAE, Uruguay

Table 3
Correlations among moral foundations.

	1	2	3	4	5	6	7
Care		0.61**	0.30**	0.23**	0.40**	0.90**	0.39**
Fairness	0.71**		0.31**	0.29**	0.40**	0.90**	0.42**
Loyalty	0.49**	0.48**		0.49**	0.42**	0.34**	0.80**
Authority	0.30**	0.33**	0.49**		0.47**	0.29**	0.80**
Purity	0.49**	0.47**	0.50**	0.50**		0.45**	0.80**
Individualizing	0.93**	0.92**	0.53**	0.34**	0.52**		0.45**
Binding	0.52**	0.53**	0.82**	0.80**	0.83**	0.57**	

Note: Upper diagonal represents correlations among WEIRD sample, lower diagonal represents correlation among non-WEIRD sample.

** $p < .001$.

MFQ on the whole sample, WEIRD sample and non-WEIRD sample separately. Models were tested for each of the two factor and five-factor solution. Afterwards, measurement invariance procedure was applied to the validated/fitted factor structure since measurement invariance requires a base model in which data fit the model well (Kline, 2011).

Measurement invariance across groups (i.e., WEIRD, non-WEIRD) was tested on the two and five-factor models individually. Though various perspectives are proposed to test measurement invariance, the most frequent approach is to start with general forms of measurement invariance followed by more specific tests (Vandenberg & Lance, 2000). In the current study, we adopted this perspective since identifying differences across groups is more likely (Kline, 2011). Therefore, first, we tested configural invariance, which implies that the same factor structure fits across groups and all the parameters are different between groups. Secondly, factor loadings invariance or metric invariance test was employed in which unstandardized factor loadings are equal across groups and all other parameters are freely estimated. Thirdly, scalar invariance, which implies an equal indicator (item) means across groups as well as equal factor loadings, was tested. Since invariance of residual variance test is highly stringent (Brown, 2014), we ended invariance tests at scalar invariance level.

3. Results

According to the results of CFAs, the five-factor model yielded good fit to the data for the whole sample, WEIRD sample, and non-WEIRD sample. However, fit indices for the two-factor model were below cut-off point for all three samples. Besides, five-factor models on each sample yielded relatively better fit as compared to the two-factor model based on the AIC and BIC criteria. As shown in Table 4, all item factor loadings on two and five-factor models were significant on each sample (all p's < 0.001).

Measurement invariance test across WEIRD and non-WEIRD samples are depicted in Table 5. First, configural invariance test was applied. Configural invariance across groups was evaluated via investigation of overall model fit. Accordingly, configural invariance model yielded good fit to data for the five-factor model, which suggests that the factor structure including number of factors and items assigned to each factor is same across groups. Results of configural invariance test for the two-factor model revealed unsatisfactory fit indices in line with the baseline models. Following configural invariance, metric invariance test was conducted to compare factor loadings of each item across samples. Results of χ2-difference-test revealed metric non-invariance across samples for the five-factor model ($\Delta\chi2(10) = 49.55$,

$p < .001$) suggesting that item loadings are not same across samples. Furthermore, the metric invariance test for two-factor model yielded similar results ($\Delta\chi2(13) = 68.47$, $p < .001$) yet fit indices were again below the acceptable level. Since metric noninvariant samples are not comparable on more strict invariance tests, we did not apply χ2-difference-test on scalar invariance.

4. Discussion

Moral foundations theory proposes five dimensions which are assumed to be universal and numerous cross-cultural research have been conducted to validate its basic premises and to explore its correlates. Considering the mixed results in the past literature on cross-cultural stability of the dimensions, it was important to validate factor structure in different cultural contexts. For this aim, we tested the measurement invariance of the MFQ in WEIRD and non-WEIRD samples. In summary, the five-factor model of MFQ revealed a good fit to the data on both WEIRD and non-WEIRD samples. Besides, the five-factor model yielded a better fit to the data as compared to the two-factor model of MFQ. Measurement invariance test across samples validated factor structure for the five-factor model, yet a comparison of samples provided metric non-invariance implying that item loadings are different across groups.

The results provided support to a five-factor solution to morality. The number of dimensions in moral convictions has been a matter of debate, not only for measurement purposes but also to identify the true number of distinct moral foundations. MFT adopts a *modular mind* perspective (see Cosmides & Tooby, 1994; Fodor, 1983) which suggests that the human mind is comprised of several modules, each of which was evolved to solve specific problems. MFT proposes that there are five such domains in morality solving different problems that might disrupt the social life throughout the evolutionary process: (1) *care/harm* foundation solves the problem of protecting the vulnerable members of the society (children, elderly, etc.) from harm; (2) *fairness/cheating* foundation has evolved the solve the problem of free riders and cheaters who might undermine the cooperative effort of human societies; (3) *loyalty/betrayal* foundation has evolved to prevent members from being disloyal to their ingroup; (4) *authority/subversion* foundation ensures that the leadership structure of the group stays intact and the group members can effectively get organized under the authority figures; and (5) *sanctity/degradation* solves the problem of protecting physical health and maintaining the spiritual integrity that binds a group together (Graham et al., 2009). However, later research challenged this categorization and argued that there were two, not five, main moral foundations. Accordingly, it was suggested that individualizing foundation (which is an aggregate of care/harm and fairness/cheating) is related to protecting the individual from being harmed and unfairly treated whereas binding foundation (which is aggregate of loyalty/betrayal, authority/subversion, and sanctity/degradation) is related to facilitating the cohesion and solidarity within the group (Van Leeuwen & Park, 2009; Wright & Baril, 2011; Yilmaz et al., 2016). As the number of foundations is directly related to the number of problems to be solved, the debate on five-factor versus two-factor solutions to morality is theoretically important to moral psychology literature. One practical way to help resolve this issue is to look at how people respond to MFQ The number of factors emerged in the patterns of responses would be a strong indicator for the number of distinct moral foundations that exist. This was what the current study aimed for and configural invariance across groups highlighted the robustness of the five-factor structure of the MFQ in the current study. In

Table 4
Factor loadings for baseline models.

	Five-factor model			Two-factor model		
	Whole	WEIRD	Non-Weird	Whole	WEIRD	Non-Weird
Whether or not someone was harmed	0.78	0.76	0.80	0.74	0.72	0.77
Whether or not someone suffered emotionally	0.70	0.66	0.66	0.66	0.65	0.65
Whether or not someone used violence	0.72	0.69	0.76	0.70	0.67	0.74
Whether or not some people were treated differently than others	0.66	0.64	0.69	0.63	0.60	0.67
Whether or not someone was denied his or her rights	0.71	0.67	0.76	0.69	0.65	0.74
Whether or not someone acted unfairly	0.65	0.63	0.74	0.62	0.59	0.71
Whether or not someone did something to betray his or her group	0.74	0.73	0.77	0.64	0.61	0.69
Whether or not the action was done by a friend or relative of yours	0.38	0.39	0.35	0.35	0.36	0.34
Whether or not someone showed a lack of loyalty	0.72	0.73	0.71	0.61	0.60	0.64
Whether or not the people involved were of the same rank or status	0.33	0.34	0.30	0.30	0.31	0.25
Whether or not someone failed to fulfill the duties of his or her role	0.56	0.56	0.57	0.53	0.52	0.54
Whether or not someone showed a lack of respect for legitimate authority	0.61	0.62	0.60	0.57	0.58	0.54
Whether or not someone did something disgusting	0.64	0.63	0.71	0.58	0.56	0.64
Whether or not someone violated standards of purity and decency	0.66	0.68	0.61	0.61	0.62	0.58
Whether or not someone did something unnatural or degrading	0.72	0.71	0.73	0.64	0.64	0.66

other words, the results supported a five-factor solution, and there are five, not two, distinct moral foundations.

There were previous attempts to establish a true factorial structure. Graham et al. (2009) initially argued for a five-factor solution. However, later research showed that the five-factor model did not have a good fit on the data (e.g., Davis et al., 2016; Yalçındağ et al., 2017). Other research suggested that it would be plausible to utilize a two-factor solution (e.g., Napier & Luguri, 2013; Van Leeuwen & Park, 2009; Wright & Baril, 2011; Yilmaz et al., 2016). A recent cross-cultural study suggested that a five-factor solution did not have a good fit to the data (Iurino & Saucier, 2019). However, we reached the opposite conclusion in our study; the cross-cultural data we examined suggested that the five-factor solution had a good fit, even in very different cultural contexts spanning the WEIRD and non-WEIRD cultures (Henrich et al., 2010).

Although our results revealed that the five-factor solution is better than the two-factor solution for WEIRD and non-WEIRD samples, it was also found that there was metric non-invariance, meaning that the factor loadings are not equivalent across groups. Despite the fact that the items measure the same moral foundations, they had varying levels of loadings to these factors. One of the implications of metric non-invariance is that the MFQ might not be a suitable tool if one aims to compare moral foundations across WEIRD and non-WEIRD cultures, because any difference could be due to the differences in loadings, rather than mean endorsement of a specific moral foundation. It should be noted that since the chi-square tests are likely to yield significant

results in large sample sizes, the possibility of metric invariance should not be ruled out. In that case, scalar invariance test provided much more conservative result by suggesting that the means of moral foundations are different across groups. Another implication of the results is related with how the endorsement of moral foundations is measured by MFQ. MFQ includes several statements that are assumed to be linked with a specific foundation (Graham et al., 2011). For example, "Whether or not someone suffered emotionally" and "Whether or not someone cared for someone weak or vulnerable" are two items tapping into the care/harm foundation. The more you think these are relevant to deciding whether something is right or wrong, the more you endorse the care/harm foundation. Our results suggested that the level of loadings of the items vary between WEIRD and non-WEIRD cultures. In other words, although the same statements tap into the same moral foundations in each case, the strength of the link between the statements and the foundations were different in WEIRD and non-WEIRD cultures. This suggests that, although there is a five-factor structure of morality in both samples, some specific situations described by the items were found to be more or less related to the underlying moral foundation in different cultures. To our knowledge, the potential differences in how much each moral foundation applies to different contexts in different cultures have not been investigated before, and future research is needed to illustrate a more complete picture of such potential cultural differences.

The current study is distinct from the previous similar investigations on several dimensions. First, unlike Graham et al. (2011), the present

Table 5
Results from measurement invariance tests.

Type of invariance	χ2 (df)	RMSEA (90% CI)	SRMR	CFI	TLI	AIC	BIC
Five Factor Model							
Whole Sample	1508.112 (80)	0.05	0.04	0.95	0.93	302,338.66	302,011.96
Weird Sample	1030.105 (80)	0.05	0.04	0.94	0.93	212,954.19	213,312.32
Non-Weird Sample	590.157 (80)	0.06	0.04	0.94	0.92	87,968.44	88,276.41
Configural invariance	1618.376 (160)	0.05	0.04	0.94	0.93	300,922.64	301,676.04
Metric invariance	1667.926 (170)	0.05	0.04	0.94	0.93	300,947.41	301,632.32
Scalar invariance	2322.131 (185)	0.06	0.05	0.92	0.91	301,629.80	302,211.97
Two Factor Model							
Whole Sample	2838.153 (89)	0.07	0.05	0.89	0.88	303,521.033	303,836.09
Weird Sample	21,076.786 (89)	0.07	0.05	0.88	0.86	214,165.38	214,464.90
Non-Weird Sample	892.299 (89)	0.07	0.05	0.91	0.89	88,304.90	88,562.47
Configural invariance	2991.316 (178)	0.07	0.05	0.89	0.87	302,470.28	303,100.39
Metric invariance	3060.786 (191)	0.07	0.05	0.89	0.88	302,493.41	303,034.49
Scalar invariance	3729.997 (206)	0.07	0.06	0.86	0.86	303,182.26	303,620.60

Note. df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = confirmatory fit index; TLI = Tucker-Lewis index; AIC = Akaike information criteria; BIC = Bayesian information criteria.

data includes responses in participants' native languages (Klein et al., 2018). In Graham et al.'s (2011) study, the data were retrieved from a website including a set of questionnaire that participants around the world responded to. However, the questionnaire was in English which limits the sample into only those who are fluent in English. Such fluency in English might indicate a familiarity with the North American culture and those eligible participants might have a cultural worldview that is more similar to the WEIRD cultures (Henrich et al., 2010). Second, the voluntary nature of participation in Graham et al.'s (2011) study could potentially lead to a "selection bias" (Wainer, 1999) as the sample included participants who were willing to spend time for a morality-related study. In our study, however, the MFT was embedded in a series of unrelated tasks and the participation was in exchange of an incentive (mostly in the form of extra grades), so there was minimal risk of selection bias.

These two limitations, in fact, were overcome in a later attempt to examine the factor structure of MFT (Iurino & Saucier, 2019). However, unlike current findings, their results did not replicate a five-factor model. One reason, as also pointed out by Iurino and Saucier (2019), could be the lack of randomization of item order in their study. In Many Labs 2 project (Klein et al., 2018), where we retrieved the data from, the order of items was randomized. Iurino and Saucier (2019) also suggested that their use of a short form of MFQ could be another reason for failure to replicate the five-factor model. However, the current study succeeded in replicating a five-factor model despite using a short form of the questionnaire.

In addition to the similarities with past studies, the present study had also an unprecedented contribution to the literature. We examined the factor structure of MFQ in both WEIRD and non-WEIRD cultures. It has been widely recognized that the vast majority of psychological findings are obtained in countries fitting the definition of WEIRD, although they make up only a small minority in the world's population (Henrich et al., 2010). Considering the potential sampling bias in the original MFQ study (Graham et al., 2011), it was crucial to investigate the structure of MFQ in non-WEIRD cultures as well. The current study provided important evidence for the universality of MFQ by showing the same five-factor solution in both WEIRD and non-WEIRD samples.

5. Potential limitations

There are also some limitations of the current study. First, Many Labs 2 project (Klein et al., 2018) included many studies, in addition to the completion of MFQ. The ordering of tasks might affect the responses, although we do not have any theoretical reason to expect so. Second, a short form of MFQ was used. Although the short-form of MFQ is widely used (e.g., Graham et al., 2011), the number of items in different studies utilizing MFQ could influence the level of reliabilities for the factors. Third, we have identified metric non-invariance, meaning that item loadings were different across groups, however explaining why metric non-invariance occurred was beyond the scope of the current research. We argue that future research should be designed in a way to overcome these limitations and enrich our knowledge regarding the validity and reliability of MFQ across different cultures.

6. Conclusion

Despite the potential limitations, the current findings have important implications for moral psychological research. First, applying a two-factor solution (individualizing and binding foundations) to the MFQ, as has been done before (e.g., Napier & Luguri, 2013; Van Leeuwen & Park, 2009; Wright & Baril, 2011; Yilmaz et al., 2016), might be simplifying the five-dimensional nature of morality. Instead, a five-factor solution is more valid and cross-culturally stable and should be preferred in future research. Second, the results are also informative regarding how many distinct moral foundations there are. The modular mind perspective in evolutionary psychology proposes that our mind is

comprised of specific modules evolved to solve specific problems (e.g., Cosmides & Tooby, 1994). Following a similar logic, MFT proposes distinct domains of morality, each of which have been evolved to solve different social problems (Graham et al., 2009). The five-factor structure of responses to MFQ suggests that there are at least five distinct domains, instead of two. As a result, the findings do not only have important implications for the measurement of moral foundations, but also the theory itself. Lastly, we have also identified that, although a five-factor solution to morality is better suited to both WEIRD and non-WEIRD nations, there are cross-cultural differences in loadings of items to each factor which necessitates future research investigating the reasons behind this difference.

References

Berniūnas, R., Dranseika, V., & Sousa, P. (2016). Are there different moral domains? Evidence from Mongolia. *Asian Journal of Social Psychology, 19*, 275–282.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* Guilford Publications.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld, & S. A. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York, NY, US: Cambridge University Press.

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.

Davies, C. L., Sibley, C. G., & Liu, J. H. (2014). Confirmatory factor analysis of the Moral Foundations Questionnaire. *Social Psychology, 45*, 431–436.

Davis, D. E., Rice, K., Van Tongeren, D. R., Hook, J. N., DeBlaere, C., Worthington, E. L., Jr., & Choe, E. (2016). The moral foundations hypothesis does not replicate well in Black samples. *Journal of Personality and Social Psychology, 110*(4), 23–30.

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology.* Cambridge, Massachusetts: MIT Press.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55–130.

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*, 1029–1046.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*, 366–385.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998–1002.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Iurino, K., & Saucier, G. (2019). Testing measurement invariance of the moral foundations questionnaire across 27 countries. *Assessment.* https://doi.org/10.1177/1073191118817916.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., … Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across sample and setting. 10.31234/osf.io/9654g.

Kline, R. B. (2011). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Kohlberg, L. (1969). *Stage and sequence: The cognitive-developmental approach to socialization.* New York: Rand McNally.

Métayer, S., & Pahlavan, F. (2014). Validation de l'adaptation française du questionnaire des principes moraux fondateurs. *Revue Internationale de Psychologie Sociale, 27*(2), 79–107.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Napier, J. L., & Luguri, J. B. (2013). Moral mind-sets: Abstract thinking increases a preference for "individualizing" over "binding" moral foundations. *Social Psychological and Personality Science, 4*(6), 754–759.

Nilsson, A., & Erlandsson, A. (2015). The Moral Foundations taxonomy: Structural validity and relation to political ideology in Sweden. *Personality and Individual Differences, 76*, 28–32.

Piaget, J. (1965). *The moral judgment of the child.* New York, NY, US: Free Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In A. Brandt, & P. Rozin (Eds.). *Morality and health* (pp. 119–169). New York, NY: Routledge.

Van Leeuwen, F., & Park, J. H. (2009). Perceptions of social dangers, moral foundations, and political orientation. *Personality and Individual Differences, 47*(3), 169–173.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70.

Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods, 4*(2), 212.

Wright, J. C., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology, 47*(5), 1007–1012.

Yalçındağ, B., Özkan, T., Cesur, S., Yilmaz, O., Tepe, B., Piyale, Z. E., ... Sunar, D. (2017). An investigation of moral foundations theory in Turkey using different measures.

Current Psychology, 1–18.

Yilmaz, O., Harma, M., Bahçekapili, H. G., & Cesur, S. (2016). Validation of the moral foundations questionnaire in Turkey and its relation to cultural schemas of individualism and collectivism. *Personality and Individual Differences, 99,* 149–154.

Zhang, Y., & Li, S. (2015). Two measures for cross-cultural research on morality: Comparison and revision. *Psychological Reports, 117,* 144–166.