

COMPARATIVE AND CROSS-CULTURAL VALIDITY OF THE MORAL  
ACTIONS QUESTIONNAIRE, A MEASURE FOR ETHICAL VIRTUE

by

KATHRYN MARY IURINO

A DISSERTATION

Presented to the Department of Psychology  
and the Graduate School of the University of Oregon  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

March 2018

DISSERTATION APPROVAL PAGE

Student: Kathryn Mary Iurino

Title: Comparative and Cross-Cultural Validity of the Moral Actions  
Questionnaire, A Measure for Ethical Virtue

This dissertation has been accepted and approved in partial fulfillment of the  
requirements for the Doctor of Philosophy degree in the Department of Psychology  
by:

Gerard Saucier	Chair
Elliot Berkman	Core Member
Sanjay Srivastava	Core Member
Kathleen Scalise	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School
----------------	--

Original approval signatures are on file with the University of Oregon Graduate  
School.

Degree awarded March 2018

© 2018 Kathryn Mary Iurino  
All rights reserved.

## DISSERTATION ABSTRACT

Kathryn Mary Iurino

Doctor of Philosophy

Department of Psychology

March 2018

Title: Comparative and Cross-Cultural Validity of the Moral Actions Questionnaire, A Measure for Ethical Virtue

Though current personality models provide a starting point for measuring ethical virtues, ethical content may not be fully captured in existing personality inventories due in part to the systematic elimination of morally-relevant trait-adjectives in early lexical studies. Further, personality dimensions relevant to measuring the ethical domain include both ethical and non-ethical content. The Moral Actions Questionnaire was designed to assess seven conceptually-distinct ethical virtues that are emphasized across cultures and philosophies. This dissertation investigates the performance of the Moral Actions Questionnaire, relative to other candidate models of ethical virtue from personality inventories. Psychometric quality, structural validity, and predictive validity for these models are evaluated in samples from five countries: Kenya, India, Hong Kong, Singapore, and the United States. Findings suggest that the Moral Actions Questionnaire aids in prediction of altruistic bravery, guilt proneness, satisfaction with life, and meaning with life across most countries. Patterns in psychometric quality and structure across countries and methods (self- and informant-report) are discussed.

## CURRICULUM VITAE

NAME OF AUTHOR: Kathryn Mary Iurino

### GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA  
Grinnell College, Grinnell, IA, USA

### DEGREES AWARDED:

Doctor of Philosophy, Psychology, 2018, University of Oregon  
Master of Science, Psychology, 2012, University of Oregon  
Bachelor of Arts, Mathematics and Psychology (with Honors), 2010,  
Grinnell College

### PROFESSIONAL EXPERIENCE:

Assessment Analyst, University Housing, University of Oregon, 2017-2018  
Research Assistant, Oregon Research Institute, 2014-2015  
Research Assistant, University of Oregon, 2011-2014  
Lab Instructor-Graduate Statistics, University of Oregon, 2016-2017  
Instructor-Culture and Mental Health, University of Oregon, 2016  
Instructor-Introduction to Statistics, University of Oregon, 2013  
Lab Instructor-Introduction to Statistics, University of Oregon, 2013-2015  
Lab Instructor-Research Methods, University of Oregon, 2015  
Teaching Assistant-Thinking, University of Oregon, 2014  
Research Assistant, Grinnell College, 2009-2010

## GRANTS, AWARDS AND HONORS:

ICPSR Program in Quantitative Methods Scholarship, University of Michigan- Ann Arbor, 2017

Marthe E. Smith Memorial Science Scholarship, University of Oregon, 2016

Beverly Fagot Dissertation Fellowship, University of Oregon, 2016

Clarence and Lucille Dunbar Scholarship, University of Oregon, 2015

Gregores Award, University of Oregon, 2014

Clarence and Lucille Dunbar Scholarship, University of Oregon, 2014

Trustee Honor Scholarship, Grinnell College, 2006-2010

The President's Volunteer Service Award, 2006

## PUBLICATIONS:

Iurino, K., Robinson, B., Christen, M., Stey, P., & Alfano, M. (forthcoming). Constructing and validating a scale of inquisitive curiosity. In Inan, Whitcomb, & Yigit (Eds). *The Moral Psychology of Curiosity*.

Iurino, K., & Saucier, G. (forthcoming). Personality structure. In Zeigler-Hill & Shackelford (Eds). *Encyclopedia of Personality and Individual Differences*.

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). Development and validation of a multi-dimensional measure of intellectual humility. PLoS ONE 12(8): e0182950. <https://doi.org/10.1371/journal.pone.0182950>

Saucier, G., Kenner, J., Iurino, K., Bou Malham, P., Chen, Z., Thalmayer, A. et al. (2015). Cross-cultural differences in a global Survey of World Views. *Journal of Cross-Cultural Psychology*, 46, 50-70.

## ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Gerard Saucier, for his feedback on this dissertation and many other projects, as well as for broadening and deepening my understanding of personality, ethical virtues, and cross-cultural research. I would also like to thank my committee, and in particular Kathleen Scalise for her tailored feedback in Item Response Theory II and advice on the IRT analysis. I would like to thank my family for their unwavering support throughout graduate school. Thanks also to Jason Vielma for being so consistently supportive for the past three years, as well as for patiently answering my LaTeX questions, and thanks to past University of Oregon graduate students who provided the LaTeX template. I am also grateful to Rain M. Baker Guardia who was a great source of intellectual support and who entertained many interesting discussions on moral philosophy and research, especially in the beginning stages of graduate school. And I am grateful to other past graduate students - specifically Amber Gayle Thalmayer, Philippe Bou Malham, Jacob Levernier, Jennifer Gómez, and Job Chen- all of whom I have been inspired and encouraged by. Finally, I would like to thank the John Templeton Foundation, who provided the funding to collect the data used for this project.

In memory of Talent Takundwa, whose brilliance and humility embodied these virtues in their most ideal form.

Also dedicated to my mother Laura M. LaFave, who is my greatest exemplar of compassion, and to my father John N. Iurino, who taught me the value of fairness, promise-keeping and perseverance.



## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION . . . . .	1
II. METHOD . . . . .	15
Participants . . . . .	15
Materials . . . . .	16
Measures of Models for Ethical Virtue . . . . .	16
Full Moral Actions Questionnaire–Self-Report Version . . . . .	16
Moral Actions Questionnaire–Informant-Report Version . . . . .	17
HEXACO Inventory . . . . .	17
Questionnaire Big Six Inventory . . . . .	17
Big Six Inventory . . . . .	18
Outcome Measures . . . . .	18
MIDUS (Midlife Development in the U.S.) 10-item Measure of Compassionate Norms . . . . .	18
Overall Perception of Contribution to the Welfare of Others . . . . .	19
Likelihood to Help a Cyclist . . . . .	19
Guilt and Shame Proneness, as Measured by the 16-item GASP . . . . .	19
Relative Immunity from Inflation Due to Self-Enhancement . . . . .	19
Likelihood of Returning Paid-in-Advance Survey . . . . .	20
Health . . . . .	21
Meaning in Life Scale (MILS) . . . . .	21
Satisfaction with Life Scale (SWLS) . . . . .	21
Analysis . . . . .	21

Chapter	Page
Psychometric Quality . . . . .	22
Structural Validity . . . . .	22
Predictive Validity . . . . .	24
III. RESULTS . . . . .	26
Data Cleaning Procedure . . . . .	26
Psychometric Quality . . . . .	27
Test-Retest Stability . . . . .	29
Item Response Theory (IRT) . . . . .	30
Structural Validity . . . . .	35
Structure in Self-Report Data . . . . .	35
Measurement Invariance Tests Across Hong Kong/Singapore, India, Kenya, and National Sample. . . . .	40
Structure in Informant-Report Data . . . . .	42
Predictive Validity . . . . .	45
Forecasting . . . . .	57
IV. DISCUSSION . . . . .	60
V. CONCLUSION . . . . .	70
APPENDICES	
A. A CLOSER LOOK AT TRUTH-TELLING AND COMPASSION . . . . .	71
B. EXPLORATORY FACTOR ANALYSIS RESULTS . . . . .	72
C. RELIABILITY OF RATERS FOR SIX MAQ DOMAINS . . . . .	76
REFERENCES CITED . . . . .	77

## LIST OF FIGURES

Figure	Page
1. Distribution of Theta estimates using MAQ inventory in self-report data by country . . . . .	32
2. Distribution of Theta estimates using 8-item Big Six H in self-report data by country . . . . .	32
3. Relationship between Theta and SE for MAQ inventory in self-report data . . . . .	33
4. Relationship between Theta and SE for Big Six H in self-report data . . . . .	33
5. Distribution of Theta estimates using MAQ inventory in informant data by country . . . . .	34
6. Relationship between Theta and SE for informant-rated MAQ inventory . . . . .	34
7. Adjusted R-squared for hierarchical models predicting each outcome . . . . .	46
8. R-squared estimates for the regression model assuming no measurement error and SEM regression model. . . . .	52

## LIST OF TABLES

Table		Page
1.	Standardized $\alpha$ for self-report scales in each country . . . . .	28
2.	Mean of inter-item correlations for self-report scales in each country . . . . .	28
3.	Variance of inter-item correlations for self-report scales in each country . . . . .	28
4.	Standardized $\alpha$ for informant-report MAQ scales in each country . . . . .	29
5.	Mean of inter-item correlations for informant-report MAQ scales in each country . . . . .	29
6.	Variance of inter-item correlations for informant-report MAQ scales in each country . . . . .	29
7.	One-year retest stability for MAQ and HEXACO-H in National Sample . . . . .	30
8.	Estimated factor correlation matrix for seven-factor model in U.S. (Community Sample) data . . . . .	36
9.	Fit of seven-, three-, and one-factor models in Community Sample . . . . .	36
10.	Fit indices for models in self-report data across countries . . . . .	39
11.	Measurement invariance tests for three-factor model (version where Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling) in self-report data . . . . .	40
12.	Measurement invariance tests for three-factor model (version where Non-Maleficence was grouped with Compassion and Beneficence) in self-report data . . . . .	41
13.	Measurement invariance tests for one-factor model in self-report data . . . . .	41
14.	Measurement invariance tests for interstitial model in self-report data . . . . .	41

Table	Page
15. Estimated factor correlation matrix in Student/Informant Sample (where latent variable covariance matrix was not positive definite) . . . . .	43
16. Fit of models in Student/Informant Sample . . . . .	43
17. Fit indices for models in informant-report data across countries . . . . .	44
18. Estimated factor correlation matrix for three-factor model in each country . . . . .	44
19. Measurement invariance tests for two-factor model in informant-report data . . . . .	44
20. Measurement invariance tests for one-factor model in informant-report data . . . . .	45
21. Zero-order correlations between ethical virtue scales and outcomes in U.S. National Sample . . . . .	53
22. Zero-order correlations between ethical virtue scales and outcomes in India Sample . . . . .	54
23. Zero-order correlations between ethical virtue scales and outcomes in Hong Kong/Singapore Sample . . . . .	55
24. Zero-order correlations between ethical virtue scales and outcomes in Kenya Sample . . . . .	56
A.25. Correlations between Truth-Telling items in self-report data . . . . .	71
A.26. Correlations between Compassion items in self-report data . . . . .	71
B.27. Results from item-level Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Student/Informant sample (N=842) (which informed structures tested in international samples)- two-factor solution . . . . .	72
B.28. Results from item-level Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Student/Informant sample (N=842) (which informed structures tested in international samples)- three-factor solution . . . . .	73
B.29. Results from Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Community sample (n=538)- two-factor solution . . . . .	74

Table	Page
B.30. Results from Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Community sample (n=538)- three-factor solution . . . . .	75
C.31. Standardized alpha values for 3 raters' ratings of 2458 IPIP items' relevance to each of the original six MAQ domains . . . . .	76

## CHAPTER I

### INTRODUCTION

The psychological study of morality has taken many forms, ranging from the study of people’s responses to hypothetical moral dilemmas and moral reasoning, to people’s actual behavior in situations relevant to morality. This dissertation takes a virtue-ethics approach to the study of morality, conceptualizing ethical virtue<sup>1</sup> as a tendency toward a particular type of moral behavior across time. As has been argued elsewhere, moral psychology could benefit from the study of moral traits as dispositions (Fleeson, Furr, Jayawickreme, Meindl, & Helzer, 2014; Hill & Roberts, 2010), rather than from the primarily situationist perspective that much of the field has adopted.

The psychological study of personality has also taken many forms, and these forms have typically under-represented the domain of moral traits. That is, in the history of personality psychology, moral content was partially excluded. The highly influential factor analyses by Goldberg (1990) supporting a Big Five structure were performed on a list of adjectives originating from earlier work by Allport and Odbert (1936), who had excluded some moral content from the variable list. Allport and Odbert (1936) had considered moral terms to be evaluative and thus not aligned with scientific standards under which objective descriptions are desired.

More recently, there have been a few notable integrations of moral character into measures of personality, most notably in the Honesty/Propriety dimension from the Big Six model of personality and the Honesty/Humility dimension in

---

<sup>1</sup>In this dissertation, the terms “ethical” and “moral” are treated as interchangeable, but since the current project focuses on moral considerations that tend to be common across ethical codes, rather than including all domains that could possibly be considered moral, the term “ethical virtues” is generally preferred.

the HEXACO personality inventory. There is evidence that six-factor models of personality are more predictive than the five-factor model (Ashton, Lee, & De Vries, 2014).

In addition, as argued by Saucier (2017), the most central dimension of personality is character— the tendency to regulate one’s selfish impulses with respect to moral norms. Evidence for the centrality of moral tendencies in personality comes from the fact that when personality is assessed using type-nouns in American-English, by far the largest dimension that emerges is primarily moral in content (Saucier, 2003), and similar results are obtained for type-nouns in the Dutch language (De Raad & Hoskens, 1990). Further, moral and social self-regulation content in personality (even if one constrains oneself to adjectives) is more ubiquitous across diverse languages from Africa, Asia and Europe than content of any Big Five factor (Saucier, Thalmayer, & Bel-Bahar, 2014).

The empirical reality that morality is a central component of the personality domain, rather than something exogenous to personality, suggests a disposition-oriented approach to understanding ethical virtues will have some benefits. Namely, the study of the ethical domain can borrow from and be integrated with the personality domain, thus helping to enrich our understanding of the broader personality domain. In addition, though some have theorized that virtuous behavior is too affected by changes in situations to be properly considered a stable attribute of an individual (Doris, 2002), Costello, Srivastava, and Saucier (2017) found that the Honesty/Propriety factor of the Big Six has a high degree of test-retest stability, similar to that of other personality traits, across both informant- and self-report methods. This suggests that measuring ethical virtues as a stable individual-differences variable is appropriate.



For these reasons, the current project takes a personality-oriented approach to assessing ethical virtues. Within personality models, there is some representation of ethical virtues already: particularly in the Honesty/Propriety dimension in the Big Six model (Saucier, 2009), and in the Honesty/Humility dimension in the HEXACO model (Lee & Ashton, 2013). In the HEXACO model, the Honesty/Humility dimension is characterized by tendencies to be manipulative versus fair and law-abiding, and also reflects varying levels of concern for status and possessions (Lee & Ashton, 2013). In the Big Six model, Honesty/Propriety reflects a similar overlapping tendency, but has less emphasis on the humility aspect, more emphasis on propriety (following social norms so as to not cause trouble). In addition, ethical content is represented in some versions of Big Five Agreeableness (including content related to compassion and generosity in Saucier and Ostendorf (1999) and Big Five Conscientiousness (including content that may be related to dependable keeping of promises, i.e., responsibility in Saucier and Ostendorf (1999), and reliability in Roberts, Bogg, Walton, Chernyshenko, and Stark (2004)). But ethical content is never the central component of either Big-Five domain; it merely appears in fractionated form in several subcomponents of these factors.

In addition, a potential limitation of these models is that the content that was included was based on factor-analytic considerations about what makes a good model of personality as a whole (including what kinds of content stays fairly orthogonal to other kinds of content), rather than a rationale-based consideration of what the basic ethical virtues are. These empirical constraints likely prevented the ethical domain from being fully represented. For measures of the Big Five model, the representation of ethical virtue is partial at best, as content was

selected to the extent that it cohered with dimensions of Agreeableness and Conscientiousness. Honesty/Humility from the HEXACO and Honesty/Propriety from the Big Six models are more exclusively dominated by ethical content than Agreeableness and Conscientiousness are, so could be thought to more directly target ethical virtue than these two factors in the Big Five. However, content related to the H domain in both these models was chosen in a way that potentially could also have had a limiting impact on the coverage of the domain of ethical virtues: presumably, in order to provide a reasonably well-fitting six factor model of personality that minimizes cross-loadings, the ethical content selected would have to be more independent from other aspects of personality. This may have excluded certain types of ethical content that are more interstitial to H and A (such as compassion and generosity), or H and C (such as promise-keeping).

Therefore, the study of the ethical domain would benefit from starting from a rationale-based consideration of what the core domains of ethical virtues are, and based on that, develop measures well-suited for measuring each domain. This is likely to result in a more complete inventory of the ethical virtues than would be obtainable given the constraints under which personality inventories are typically constructed. Predictive validity and reliability of this rationally-derived model can be assessed relative to alternative models of ethical virtues that have arisen more incidentally, in the context of optimizing the factor structure of the entire personality domain.

With these considerations in mind, Iurino (2012) created preliminary measures of six domains of ethical virtue. The six domains were selected so as to be consistent with a “minimalist” definition of morality (Bok, 1996), that is, including “common-denominator” content that is considered relevant to morality

across diverse philosophies and cultures. The goal was to develop measures that will be relevant to measuring individual differences in behavior across diverse populations, with the goal of contributing to an understanding of what is universal about human morality rather than what is culturally-specific. The six “common-denominator” domains originally selected were: beneficence, non-maleficence, fairness, promise-keeping, truth-telling, and gratitude/reciprocity. These six domains will be described next.

First, doing good for others (beneficence) and not harming others (non-maleficence) are universal human values (Schwartz & Bardi, 2001), emphasized in moral philosophy (Bok, 1996; Nagel, 1991; Ross, 2003), in religious and non-religious doctrines, and across many different cultures (Graham et al., 2011). More specifically, Schwartz and Bardi (2001) found that Benevolence values— defined as those values that have to do with contributing to the welfare of people with whom one interacts with frequently (helpful, honest, forgiving, loyalty, responsible)— were ranked as the most important values across a wide range of countries. Self-Direction values (creativity, freedom, independent, curious, and choosing one’s own goals) and Universalism values (broad-minded, wisdom, social justice, equality, a world at peace, a world of beauty, unity with nature, protecting the environment) tied as second most important across countries.

Second, distinct from beneficence but a natural complement to it is non-maleficence. Ross (2003) distinguishes between two morally significant classes of duties: duties that exist because there are other beings who we can improve the condition of (beneficence) and duties not to injure others (non-maleficence). Bok (1996) suggests that positive duties of mutual support and reciprocity, as well as negative duties to refrain from harm, are among the fundamental human values

that are recognized as morally significant across diverse cultures. In addition, Nagel (1991) describes duties of non-maleficence arising from the general rights that all people have to be treated in certain ways- for example, to be free from assault and coercion.

Third, a sense of fairness is likely a human universal, violations of which are perceived even by some nonhuman primates (Brosnan & De Waal, 2003) and emphasized by philosophers (i.e., Bok, 1996; Ross, 2003). According to Bok (1996), all societies have some rules or procedures intended to prevent injustice in some basic manifestations- for example, by listening to both sides in a dispute. Ross (2003) also identifies duties of justice, defining them as ensuring that pleasure or happiness is distributed in a way that is aligned with the merit of persons involved.

Fourth (and fifth), promise-keeping and truth-telling behaviors are essential to creating higher levels of trust between members of a society and groups, which is integral to maintaining a high degree of cooperation. Promise-keeping and honesty are emphasized by philosophers as important domains of moral behavior (i.e., Kant, 2001; Nagel, 1991; Ross, 2003). Further, deceit is consistently rejected in religious doctrines (e.g., Ten Commandments, Buddhist doctrine, Confucian doctrine) and secular doctrines (Bok, 1996), and "...injunctions against violence, deceit, and betrayal are found in most societies" (Bok, 1996, p. 15).

Sixth, behaviors relating to gratitude and reciprocity are also emphasized across cultures (Bok, 1996). In addition, Ross (2003) suggests behaviors relating to gratitude and reciprocity (the tendency to show thanks and return favors) are important moral duties. Some speculate that strong reciprocity- a tendency to cooperate with those who cooperate and punish those who fail to cooperate at

a personal cost— may have contributed significantly to the evolution of human altruism (Bowles & Gintis, 2011). Nagel (1991) also describes a different category of obligations arising from the special relationships we have with others— such as to one’s parents, the place where one works, or to one’s community. According to Nagel, these obligations are separate from the general obligation arising from the rights of everyone to be treated in a certain way, as they are incurred by one’s reciprocal relation to specific others. Thus, multiple theorists converge on the idea that a distinct category of value guiding action involves a set of moral duties conferred by reciprocal relationships.

Later, a seventh domain— compassion— was added, inspired by interviews done with selected members of the Eugene-Springfield Community Sample in the summer of 2014<sup>2</sup>. Though suggested from interviews with laypersons, this domain also relates to the moral-philosophy literature. From the perspective of moral sentimentalism (e.g., Hume), social emotions such as sympathy and compassion are prime sources of moral knowledge and motivation. Compassion not only informs knowledge about moral principles (Nussbaum, 1996), but is also considered a key ethical virtue by some philosophers. Schopenhauer (2014) considers compassion the most fundamental of the ethical virtues, arguing that it is the sole basis for voluntary acts of justice and genuine loving-kindness. Across disciplines, compassion is defined as having both an affective and behavioral component— Schopenhauer (2014) defines compassion as the participation in the sufferings of another, independent from any self-interested motives, resulting in an effort to prevent or remove them, and in an empirical review, Goetz, Keltner, and

---

<sup>2</sup>An eighth domain of ethical virtue (tolerance) was also considered, also inspired by laypersons’ conceptions of ethical virtue in the interviews conducted in the summer of 2014. However, the initial Tolerance subscale showed too high correlations with all the other seven domains, indicating at least the way it was measured, it was not an independent domain.

Simon-Thomas (2010) define compassion as the feeling that arises when witnessing another's suffering, and the subsequent motivation to help ease that suffering.

Compassion is distinct from beneficence and non-maleficence due to the fact that its definition contains affective and not just behavioral components.

The original item pool for each of the original six domains was created by having four raters (the author and members of the author's research team) independently evaluate the extent to which each item in the International Personality Item Pool (2458 items total) fell into the six categories of moral considerations on a continuous rating scale. The options were -2 (indicating the item related strongly to the negative pole of the domain), -1 (indicating the item related moderately to the negative pole of the domain), 0 (indicating the item did not relate the domain), 1 (indicating the item related moderately to the positive pole of the domain), or 2 (indicating the item related strongly to the positive pole of the domain). The International Personality Item Pool (IPIP) is a rather comprehensive personality-item pool, developed to facilitate evaluation of the comparative validity of a wide range of personality inventories in the interest of developing better validated personality inventories that would be publicly available (see [ipip.ori.org](http://ipip.ori.org)).

Two of the raters evaluated all six domains, and two of the raters rated only three of the six, thus the result was three raters for each domain. Overall reliability for the three raters for each domain was good: Cronbach's alpha based on standardized items ranged from a low of .634 for Fairness to .733 for Beneficence (see Appendix C for full list of reliabilities). Unstandardized alpha was much lower for Promise-Keeping, Truth-Telling, and Gratitude/Reciprocity, but this was because one of the raters had much more variance in his/her ratings of

these domains, suggesting a more inclusive interpretation of each of these concepts. Standardized and unstandardized alpha were very similar for the other domains. The 127 items rated as having the very highest extent of relevance to at least one of the six categories were considered in making one subscale of moral behavior for each of the domains. Then, using IPIP data from multiple time points, beginning from 1993, these scales were further refined, ending with six subscales with 5-6 items per scale. The scales were reduced to have a fairly low number of items from an early stage given the practical expectation that data-gathering opportunities (inclusion in large questionnaires that were measuring many other constructs each fairly briefly) would require brief scales.

Using data from the Life and Time Study (Wave 3), Iurino (2012) found that an abbreviated informant-version of this measure of moral behavior was related to self-reports of elements of what could be considered a “virtue-conducive mindset.” The informant-rated moral behavior measure was found to be positively related to Schwartz values related to helping and being honest (Schwartz, 2009), negatively related to unmitigated self-interest (Saucier, 2000), and to a lesser degree, positively related to religiosity (Iurino, 2012). The informant-version was abbreviated even more than other versions due to survey-length limitations: in this study, informants were compensated with a drawing, and a longer survey would have negatively impacted their motivation to participate.

In more recent work, the MAQ scales were refined using data from U.S. samples: a 2015 administration to the Eugene-Springfield Community Sample, (described in more detail below), Wave 1 from the Personality and Priorities Study (the National Sample, described in more detail below), and University of Oregon (UO) students’ ratings of someone they knew. For the UO students’

ratings of someone they knew, items were administered in the same format as the Student/Informant Sample, described in more detail below. For this initial scale refinement, General Survey data from Winter and Spring 2015 was used ( $N=406$ ;  $M_{\text{age}}=19.9$ ,  $SD_{\text{age}}=3.9$ ; 61.3% female).

The following characteristics were considered in selecting items, resulting in a final measure with 28 items (2 forward-keyed and 2 reverse-keyed items for each of the 7 domains):

1. Psychometric characteristics across all three samples (strength of item-total correlations and unidimensionality as assessed both by the standard deviation of each item's correlations with other items, and the number of correlations with other items that fell within .05 of the overall average inter-item correlation).
2. Structural validity, as assessed in the Community Sample and the National Sample: using SEM, items with high levels of cross-loadings within a seven-factor structure were identified. According to this criterion, preference was given to those items that did not show a strong tendency to cross-load on multiple factors.
3. Construct validity, evaluated in the Community Sample and the National Sample, by taking into account the correlations of items with the Guilt and Shame Proneness Scale (GASP; all 16 items aggregated), MIDUS measure of compassionate norms (all 10 items aggregated), Likelihood to help a cyclist (single item), and a single MIDUS item evaluating one's perception of one's overall contribution to the welfare of others. The overall strength of the



correlation with these validity indicators was determined by adding up the raw  $R^2$  values.<sup>3</sup>

4. As an independent indicator of validity from 3 above, the correlation of each item with the informant-rated composite of that same domain was evaluated.<sup>4</sup>
5. Divergent validity evaluated in the Community Sample and the National Sample—specifically, evidence that the item was measuring ethical virtue rather than socially desirable responding—evaluated based on how much correlation of the item with the item “I see myself as someone who has strong moral/ethical qualities” stayed the same when other desirable characteristics (e.g., attractiveness, likability, intelligence) were controlled for.

To arrive at a final set of items, each of these criteria was weighted one unit per sample that it was evaluated in, except for the informant-report criteria (#4 above), which was double-weighted. This resulted in a final aggregate ranking for each item reflecting its overall strength on these criteria, upon which I based the final selection of items, under the constraint that the resulting set would contain a balance of reverse- and forward-keyed items. Where the overall ranking was approximately equal, items were preferred that would increase the breadth of content representation in the scale, or the spread of item means (as a proxy for item difficulty).

---

<sup>3</sup>In future scale development, it would be useful instead to use multiple regression to evaluate the strength of the relationship between each of these items with these indicators of validity, so that the criterion does not favor the overlap between these variables.

<sup>4</sup>Though items with the same content would be expected to have higher correlations, this was not thought to be a problem, since preference should be given to items that are also in the informant-report.

The main question that this dissertation addressed was: **In what ways is the Moral Actions Questionnaire an improvement over existing measures of the ethical domain, and in what ways is it inferior to existing measures?** The psychometric quality and validity of the MAQ scales were evaluated by looking at these items' performance in comparison with other already-established measures of the ethical domain, across two different methods (self- and informant-report), and across samples from five different countries (U.S., Kenya, Hong Kong, Singapore, and India). The measures of the ethical domain were evaluated with respect to three main categories: psychometric quality, structure, and predictive validity. In the psychometric analyses, the measures of the ethical domain evaluated included the MAQ, the H dimension of the HEXACO and the H dimension of the Big Six. Big-Five models were not emphasized for the psychometric analyses because neither Agreeableness nor Conscientiousness, though each partly captures some content that could be considered in the domain of ethical virtues, is centered on the ethical virtues, and many such virtues seem to fall outside of or between them. However, Big Five Agreeableness and Conscientiousness were included (in addition to the MAQ, H from HEXACO, and H from the Big Six) in the predictive validity analyses, as it is useful to compare the extent to which all these scales predict relevant outcomes.

Investigation into the structure of the MAQ was guided by a couple of theoretical considerations. Though the MAQ is composed of seven domains with distinct moral content, it was expected that a more parsimonious factor structure would replicate better across countries. One possibility considered was that a theoretical distinction proposed by Findlay (1961) between domains centered on hortatory duties (defined as duties that are praised if done, but are

not penalized if one fails to act in accordance with them) and domains centered on minatory duties (defined as rules of strict obligation, deviation from which are penalized) would best fit the data. Of the seven domains measured in the MAQ, two could be considered hortatory (beneficence and compassion) and four could be considered minatory (promise-keeping, truth-telling, fairness, and non-maleficence). If Findlay's conception were accurate, it is possible that two broad factors would underlie responses to items in these six domains.

Another distinction proposed by Elster (2007) is between quasi-moral norms (that are conditional on the behavior of others) and moral norms (that are supposed to be unconditionally followed regardless of how others are behaving). The only quasi-moral ethical virtue assessed by the MAQ is gratitude/reciprocity, as it is the only virtue that is defined by behavior performed in response to the behavior of others, in the context of specific interpersonal relationships. All of the other ethical virtues in the MAQ do not require specific action on the part of others to be relevant. If Elster's theory were correct, one might expect the gratitude/reciprocity domain to form its own factor.

In summary, based on the above considerations, I planned to test a three-factor structure in addition to the seven-factor structure, corresponding to the domain of hortatory ethical virtue (compassion and beneficence), minatory ethical virtue (promise-keeping, truth-telling, fairness, and non-maleficence), and quasi-moral ethical virtue (gratitude/reciprocity). In addition, I planned to test a one-factor structure, since the moral domain has been modeled in other personality inventories as a single factor. I also planned to test other factor structures based on exploratory analyses done in a sample reserved for the purpose of developing

models. More details on the samples used for development and cross-validation of structural models for the MAQ are described in the Method section.

As discussed previously, these measures were also evaluated for their ability to predict a variety of outcomes that would provide support for these measures' validity (these outcomes included guilt and shame proneness, compassionate norms, altruistic bravery, overall evaluation of contribution to the well-being of others, and an objective behavior related to dutifulness). The main question of interest was whether the seven subscales of the MAQ are redundant with already-existing models, or whether any of these subscales predict content that is not covered by these models. In addition, the virtues centering on helping others (beneficence and compassion) were expected to predict the validity outcomes related to helping others (welfare contribution, compassionate norms, altruistic bravery) more strongly than the rule-focused ethical virtues such as promise-keeping and fairness, while the rule-focused ethical virtues were expected to predict the objective behavioral criterion of dutifulness more strongly.

In addition to predictive validity, the measures of ethical virtue were evaluated for their prediction of important outcomes related to well-being (health, meaning in life, and satisfaction with life). Though there are some philosophical theories that suggest that the ethically virtuous should have more meaningful lives, and higher levels of well-being (e.g., Socrates, Aristotle), there is not clear empirical support for this notion, and an argument could be made for the opposite relation (more virtuous people might be more likely to be taken advantage of, and thus might have lower well-being). However, it would be an important result if a relationship were found between ethical virtues and well-being, so these relationships were examined for exploratory purposes.

## CHAPTER II

### METHOD

#### **Participants**

There were four samples of participants.

*Community Sample.* Members of the Eugene-Springfield Community Sample were invited to participate in 2015. In the 2015 administration,  $N=538$ , \$30 compensation,  $M_{\text{age}}=70$ , 98% were Non-Hispanic White, 54% ( $N=295$ ) were female.

*National Sample.* Personality and Priorities Study National Sample. Participants for this sample were recruited partly by self-referral of Life and Time Study participants when informed of the possibility of joining another study, and partly from self-referral after completing a short screening questionnaire on Amazon Mechanical Turk. A few also came from Craigslist ads posted strategically in locales where a more diverse pool of applicants might be recruited. Wave 1 was collected in 2015; Wave 2 a year later, in 2016.  $N=608$  for Wave 1;  $N=499$  for Wave 2.  $M_{\text{age}}=35$  for Wave 1;  $M_{\text{age}}=36$  for Wave 2. In Waves 1 and 2, participants were 66% female. In Wave 1, participants were 68% Non-Hispanic White, 10% African-American, 10% Hispanic, 8% Asian/Asian-American, 2% Native American, and 1% Native Hawaiian or Pacific Islander. In Wave 2, participants were 69% Non-Hispanic White, 11% African-American, 9% Hispanic, 9% Asian/Asian-American, 1% Native American, and 1% Native Hawaiian or Pacific Islander.

*Student/Informant Sample.* This sample was constituted by Human Subjects Pool participants at the University of Oregon who completed the General Survey, from the terms: Summer and Fall 2015, Spring and Fall 2016, and Winter 2017 combined ( $N=1294$ ). For this sample, undergraduates from the University of

Oregon ( $M_{\text{age}}=19.7$ ,  $SD_{\text{age}}=2.94$ ; 68.2% female) were asked to rate someone they knew on expanded versions of the seven MAQ scales, intermixed with other personality items. Approximately a quarter of participants were asked to choose a family member, a quarter were asked to choose a friend, a quarter were asked to choose a co-worker, and a quarter were asked to choose a romantic partner. Participants were also asked how long they knew person they were rating, how similar they perceived themselves to be to the person they were rating, and demographic information about the person they were rating (age, gender, religiosity).

*International Sample.* Personality and Priorities Study, International Sample (N=910). People from four countries constituted the International Sample: India, N=422 (31% female,  $M_{\text{age}}=31$ ), Singapore (N=210, 53% female,  $M_{\text{age}}=38$ ), Hong Kong (N=59, 61% female,  $M_{\text{age}}=35$ ), and Kenya (N=219, 32% female,  $M_{\text{age}}=24$ ). All questionnaires were administered in English. The locales constituting the International Sample were selected in part due to their ability to handle English questionnaires, avoiding the need for translated questionnaires.

## **Materials**

### **Measures of Models for Ethical Virtue.**

*Full Moral Actions Questionnaire–Self-Report Version.* To assess ethical virtue, participants in the Community Sample, the National Sample (Wave 1), and the Student/Informant Sample were given a self-report version of the long form (33 items, 5-6 items per the six scales) of the MAQ developed in 2012. Participants in the National Sample (Wave 2) and the International Sample were administered the refined 28-item version of the MAQ which included the additional seventh Compassion scale.

***Moral Actions Questionnaire–Informant-Report Version.*** Two items from each of the seven domains were also administered to informants in the National Sample. These were exactly the same as the self-report versions, except for minor changes in wording necessary to fit the item for informant-report. For example, the self-report item “I see myself as someone who breaks my promises” became “I see this person as someone who breaks his/her promises” for informant-report. The 28-item version described above was also given in informant-report form to the International Sample, with the instructions that participants think of someone they admire to evaluate on these items.

***HEXACO Inventory.*** The HEXACO personality inventory (Ashton & Lee, 2004) captures a six-factor model and is increasingly widely used. The 60-item HEXACO inventory was administered for the Community Sample, a shortened 40-item version was administered to the National Sample in Wave 1, and 28 items of the HEXACO were administered to the National Sample in Wave 2. A 20-item version was also administered to the International Sample. The primary dimension of interest for this study was Honesty/Humility, as it provides a candidate model for ethical virtue. The full H dimension was administered even in the shortened versions, while items relevant to the other personality dimensions were reduced to partial versions. The other personality dimensions are less central to the domain of ethical virtue so were not included in the current analysis.

***Questionnaire Big Six Inventory.*** The 36-item version of the Questionnaire Big Six (Thalmayer, Saucier, & Eigenhuis, 2011) in meta-perception format was administered to the Community Sample, and a 30-item version was administered to Wave 1 of the National Sample. The QB6 has been studied across more than 25 countries, and was found to have approximately the same degree of

fit in tests of measurement invariance across countries as the Big Five and other contemporary inventories have across samples within the United States (Saucier et al., 2014). The “meta-perception” format— where respondents are asked to describe how they believe others see them (Carlson, Vazire, & Furr, 2011)— may help eliminate some of the self-enhancement bias that could potentially occur (particularly for valued traits) and may increase agreement between self- and informant-reports. The H scale of the QB6 in the meta-perception format provides a candidate model for ethical virtue.

***Big Six Inventory.*** The 8-item Honesty/Propriety dimension appended to the Big Five Inventory is another candidate model for ethical virtue. Items were selected for this dimension so as to be relatively independent from the other five dimensions in the full Big Five (John & Srivastava, 1999). Participants were administered the full Big Six (equivalent to the Big Five plus the 8-item Honesty/Propriety dimension) in the Community Sample, the National Sample, and the International Sample.

#### **Outcome Measures.**

***MIDUS (Midlife Development in the U.S.) 10-item Measure of Compassionate Norms.*** Described in Marks and Song (2009). This scale includes items with the stem: “Here is a list of hypothetical situations. Please rate how much obligation you would feel if the following situations happened to you.” The items include obligations to one’s family (e.g. “To drop your plans when your spouse seems very troubled”), friends (e.g. “To take your friend into your home who could not afford to live alone”), and the wider community (e.g. “To collect contributions for heart or cancer research if asked to do so”).



***Overall Perception of Contribution to the Welfare of Others.***

Additionally from the MIDUS scale, there was a single item assessing global contribution to the welfare of others on a 0 to 10 scale: “How would you rate your contribution to the well-being and welfare of others?” This outcome variable was evaluated separately from the aggregated measure of compassionate norms described above. Since it is a single item, it was expected to be more difficult to predict than the aggregate scale scores.

***Likelihood to Help a Cyclist.*** This was assessed using a single item that asked about a person’s likelihood to jump out and help a cyclist who had been hit by a car up ahead; assesses propensity for brave action. Since it is a single item, it was expected to be more difficult to predict than the aggregate scale scores.

***Guilt and Shame Proneness, as Measured by the 16-item GASP.*** Cohen, Wolf, Panter, and Insko (2011). This measure contains 8 items measuring guilt proneness (i.e., the propensity to engage in negative behavior evaluations and repair behaviors in response to hypothetical private transgressions) and 8 items measuring shame proneness (i.e., the propensity to evaluate oneself negatively and engage in withdrawal behaviors in response to hypothetical public transgressions).

***Relative Immunity from Inflation Due to Self-Enhancement.***

This was evaluated by items asking people to rate the extent to which they think they have a number of desirable but non-ethical qualities, including whether they are someone who “is likable,” “is physically attractive,” “is competent and effective,” “is smart and intelligent,” and “is a prominent important person.” One explanation for a high score on a self-reported virtue assessment would be

that truly high levels of that virtue caused the high score. A plausible alternative explanation is that a more general self-enhancement bias inflated the score.

Assuming that there is a general tendency to “self-enhance” that does not vary depending upon the domain of enhancement, then the extent to which the ethical virtue’s correlation with “Has strong moral/ethical qualities” stays the same after controlling for all these other non-moral desirable characteristics indicates to what extent the ethical virtue is entirely unaffected by general self-enhancement tendencies (i.e., independent from non-moral desirable qualities).

***Likelihood of Returning Paid-in-Advance Survey.*** This is an objective behavioral measure which was used as a validation criterion. Those participants for whom mailing addresses were known were offered the chance to complete a short (5-10 minute survey) for a small \$5 compensation, which arrived with the survey. Therefore, the participants had the option to take the money and not complete the questionnaire, or to complete and return the questionnaire to us. The choice of whether to engage in reciprocity with the research team is the behavior of interest, not the content of the survey responses. Of course, there are other non-moral reasons why someone might not return a questionnaire (such as forgetfulness, being easily distracted, lack of interest, having more important things to do) but one component is dutifulness: those high in ethical virtue should be more dutiful and return the survey, and do so quickly. The strength of this variable was hoped to be increased by our asking, in the short printed questionnaire, whether the individual had a lot of leisure time at the moment, or found the questionnaire content interesting. Among people who returned the survey, another criterion was the time it took to return the survey; one would expect that people who returned the survey more quickly, controlling for how busy

they were and interest in the survey's content, to be more virtuous than people who returned the survey less quickly.

***Health.*** This is a one-item measure where people are asked to rate their perception of their health on a 5-point scale from Excellent to Poor, as in DeSalvo, Fan, McDonell, and Fihn (2005). This measure was administered to the Community Sample, the National Sample, and the International Sample.

***Meaning in Life Scale (MILS).*** This is an eight-item measure from Schnell (2009). This measure was administered to the Community Sample, the National Sample, and the International Sample.

***Satisfaction with Life Scale (SWLS).*** This is a five-item measure; see Pavot and Diener (1993). This measure was administered to the Community Sample, the National Sample, and the International Sample.

## **Analysis**

Preliminary analyses included a side-by-side comparison of the scale characteristics of the MAQ and other candidate models of ethical virtue, from the Community Sample, National Sample, and the Student/Informant Sample. Importantly, since the final items for these scales were chosen based on their performance in two of these samples in past work, the MAQ scales' performance was expected to be inflated by over-fitting to chance characteristics of these samples. However, evaluating the MAQ scales in these same samples provided a good baseline for comparison with their performance in other samples. As a more fair comparison, the seven four-item MAQ scales were cross-validated in the International Sample to see how they performed in non-U.S. populations relative to other measures of ethical virtue, including both self-report versions and informant-report versions identical in content.

The main criteria that the scales were evaluated on in the International Sample– psychometric quality, structural validity, and predictive validity– are described in more detail below. All analyses were completed using R (R Core Team, 2017).

**Psychometric Quality.** Psychometric properties of the MAQ (both self- and informant-report versions) were evaluated separately in a subsample from each country (internal consistency was evaluated by examining the average item-total correlation and/or alpha (as calculated using the psych package in R (Revelle, 2017), unidimensionality was evaluated examining the variance of the inter-item correlations, which approaches zero as unidimensionality increases). For comparison, H from HEXACO and H from the Big Six were also evaluated on these criteria.

In order to determine how well the MAQ scales discriminated across all levels of the latent trait, a one-factor Rasch (fixed item slopes) Partial Credit Model was fit to the 28 MAQ items using the TAM package in R (Robitzsch, Kiefer, & Wu, 2018). For comparison, the H measure from the Big Six (only available in these data in self-report) was also evaluated on this criterion.

Test-retest stability was also evaluated for the MAQ in the National Sample by looking at the relationship between MAQ responses in Wave 1 and MAQ responses in Wave 2. For comparison, test-retest stability was also examined across these two time-points for the H measure from HEXACO.

**Structural Validity.** One test of structural validity is convergence of structure across methods– here, self- and informant-report methods. Structural validity was examined by looking at the fit of Structural Equation Models for the seven-factor MAQ self-report version, as well as the seven-factor informant-report

version. The lavaan package in R was used to fit all structural equation models (Rosseel, 2012).

As described in the introduction, alternative models simpler than the seven-factor model were developed in the U.S. Community Sample (for self-report data) and in the Student/Informant Sample (for informant-report data). The self-report models were then cross-validated in Wave 1 of the National Sample and in the international samples. The more parsimonious models were based in part on the theoretically-guided distinction between hortatory, minatory, and quasi-moral ethical virtues described in the introduction, and were also informed by results from an EFA performed in the U.S. Community Sample (see Appendix B for results). The pattern of factor correlations for the seven-factor CFA models in the U.S. Community Sample was also examined to guide ideas about structure.

Since it is possible that the informant-data has a different underlying structure than the self-report data, more parsimonious structures for the informant data were developed separately by performing EFAs on informant data (from the Student/Informant Sample), and also by looking at the pattern of factor correlations in the Student/Informant Sample for the seven-factor CFA. These structures were also cross-validated in the informant data from the international samples.

Measurement invariance for the self-report and informant-report MAQ models was evaluated across countries. One U.S. sample (Wave 1 of the National Sample for self-report data and the Student/Informant Sample for informant-report data) was also included in the measurement invariance analysis for each data type.

**Predictive Validity.** Validity for the self-report version of the MAQ was evaluated by checking concurrent zero-order correlations of each MAQ scale with the following outcome measures expected to be related to ethical virtue: the MIDUS measure of compassionate norms, likelihood of helping a cyclist, overall perception of contribution to the welfare of others, guilt-proneness and shame-proneness as measured by GASP, and agreement with the item “I am someone who is an ethical person” when variability due to other desirable characteristics was removed. In addition, concurrent zero-order correlations of each MAQ scale with the following well-being measures were evaluated: the single-item evaluation of health status, the Meaning in Life Scale, and the Satisfaction with Life Scale.

Participants in the international samples were also asked to choose one person they admired, and to rate that person on the MAQ. Another indication of validity for the self-report version of the MAQ was how ethically virtuous the person who was chosen was. The expectation was that, reflecting their ideals, more ethically virtuous people would be more likely to choose an ethically virtuous person as someone they admire. Further, the MAQ’s performance was compared with the performance of alternative models of ethical virtue on the same criteria (H from HEXACO, H from Big Six, A and C from Big Five).

The above was supplemented with a hierarchical regression. Overall, the goal of the hierarchical regressions was to examine whether there is content that the other candidate models of ethical virtue, gleaned from already-existing personality inventories, fail to predict, which may be better predicted by a fuller measure of ethical virtue. Already-existing measures for ethical virtue were entered first (A and C from the Big Five, H from Big Six, and the four H facets from HEXACO), and it was observed whether the MAQ scales predicted the outcomes

above and beyond what these measures predicted. The regression models were adjusted for shrinkage in order to adjust for the advantage given to the MAQ by the fact that it has more subscales than any other measure. Since one cannot assume factorial invariance for all these predictors and outcomes (measurement invariance was only evaluated for the MAQ constructs) regression analyses were conducted within each country separately. Without factorial invariance, one cannot assume that relationships found between constructs are comparable across countries, so it would not make sense to examine the relationships in aggregated data.

Forecasting power of the MAQ was examined by looking at the relationship between MAQ scores in Wave 1 of the National Sample and an objective measure of ethical virtue evaluated in Wave 2 of the National Sample which was not used to select items for each domain— Time to return a paid-in-advance survey. The first criterion was whether people higher in ethical virtue were more likely to return the survey at all. The second criterion was evaluated among people who returned the survey, using a regression analysis to see whether the time it took to return the survey (as indicated by postmark date) predicted ethical virtue, controlling for how much people enjoyed filling out the survey, and how busy people reported they were. Controlling for these other factors that may potentially affect time-to-return for the survey was expected to give a clearer picture of whether the sense of dutifulness that compels people to return the survey is uniquely related to ethical virtue as we have measured it. Since time-to-return for the survey was positively skewed (most returns coming soon after receipt of the survey), it was transformed with  $\log_{10}$  function to reduce the extent of the skew.

## CHAPTER III

### RESULTS

#### **Data Cleaning Procedure**

Beneath some threshold for completion time, respondents might be assumed to be responding impossibly fast, perhaps instead reflecting careless responding. Using a one-second-per-item threshold, from the Kenya, India, Hong Kong, and Singapore samples, cases were removed where participants took less than 109 seconds to complete the 109 self-report BFI items. Since the National Sample had 118 BFI items to complete, cases were eliminated where participants took less than 118 seconds to complete the 118 items. This criteria resulted in 11 cases in Hong Kong being eliminated (17% of original sample), 116 cases in India being eliminated (27% of the original sample), 43 cases in Kenya being eliminated (17% of original sample), 61 cases in Singapore being eliminated (29% of original sample), and 72 cases in the National Sample being eliminated (12% of original sample). Cases were also eliminated which had a standard deviation of less than .5 in their BFI responses, which resulted in an additional 5 cases from India but no more cases eliminated from any other country. After eliminating these cases, across all five countries the final sample had 1256 participants (N=536 in the National Sample, N=149 in Singapore Sample, N=205 in the Kenya Sample, N=312 in the India Sample, N=54 in the Hong Kong Sample). In these 1256 remaining participants, the majority of participants (99%) had no missing data in the 108 BFI items that were shared across the five countries; 2 participants had 2 missing data points out of the relevant 108 BFI items, 1 participant had two missing responses, and 7 participants had 1 missing response.



Since the majority of participants in the Singapore Sample indicated that they were of Chinese ethnicity (79.5% or 167 out of 210 for the full sample, and 75.8% or 113 out of 149 for the cleaned data) and the Hong Kong Sample was so small (and 81% indicated they were of Chinese ethnicity), the Singapore Sample was combined with the Hong Kong Sample for all analyses.

### **Psychometric Quality**

In general, alpha was low for the MAQ scales in self-report data, but higher for informant-report data (see Tables 1 and 4). In particular, Beneficence, Non-Maleficence and Compassion would benefit from having more items to increase alpha. Though alpha values for the MAQ scales were lower than the Big Six-H and HEXACO-H, the mean inter-item correlations tended to be higher or similar as those for Big Six-H and HEXACO-H in self-report data (see Table 2). Exceptions were Compassion and Truth-Telling, which would benefit from having a different set of items with higher internal consistency. The MAQ scales with the best reliability across countries and methods (self- and informant-report) were Promise-Keeping and Fairness (see Tables 1, 2, 4, and 5). However, especially in the Kenya and India informant samples, these scales were more multi-dimensional than is desirable (see Tables 3 and 6).

HEXACO- H had the best internal consistency across countries in self-report data, but was weaker on unidimensionality across countries. MAQ-Compassion and MAQ-Gratitude/Reciprocity were also more multidimensional than is desirable, in both self- and informant-report data.

Table 1. Standardized  $\alpha$  for self-report scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat	BSix-H	HEX-H	Meta-H
U.S. (Com)	.549	.648	.504	.578	.485	.680	.668	.673	<b>.706</b>	.636
U.S. (Nat)	.594	<b>.770</b>	.659	.631	.535	<b>.753</b>	<b>.713</b>	.623	<b>.763</b>	
Kenya	.369	.378	.551	.687	.094	.613	.603	.560	<b>.723</b>	
Sing/HK	.320	.567	.464	.646	.348	.680	.563	.560	.694	
India	.330	.417	.429	.547	.138	.656	.516	.571	.643	

*Note.* In bold are coefficients that are greater than .7, a common threshold for acceptability

Table 2. Mean of inter-item correlations for self-report scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat	BSix-H	HEX-H	Meta-H
U.S. (Com)	.233	.315	.203	.255	.190	.347	.335	.205	.193	.225
U.S. (Nat)	.268	.455	.326	.300	.224	.432	.453	.171	.244	
Kenya	.128	.132	.235	.354	.025	.284	.276	.137	.207	
Sing/HK	.105	.247	.178	.314	.118	.347	.244	.137	.185	
India	.110	.152	.158	.232	.039	.323	.211	.142	.153	

Table 3. Variance of inter-item correlations for self-report scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat	BSix-H	HEX-H	Meta-H
U.S. (Com)	<b>.003</b>	<b>.003</b>	<b>.009</b>	<b>.006</b>	<b>.006</b>	<b>.005</b>	.018	<b>.006</b>	.010	<b>.005</b>
U.S. (Nat)	<b>.006</b>	<b>.004</b>	<b>.007</b>	<b>.009</b>	<b>.006</b>	<b>.009</b>	.010	.017	.015	
Kenya	.021	.012	.010	.011	.012	.035	.021	.010	.027	
Sing/HK	.011	.012	<b>.009</b>	.017	.017	.023	.059	.015	.016	
India	.025	.015	.015	.023	.010	.012	.019	.023	.020	

*Note.* In bold are coefficients that are less than .010

Table 4. Standardized  $\alpha$  for informant-report MAQ scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat
U.S. (Student/Informant)	.660	<b>.700</b>	<b>.720</b>	<b>.752</b>	.371	<b>.782</b>	<b>.793</b>
Kenya (Informant)	.513	.344	.538	<b>.709</b>	.180	.692	.621
Sing/HK (Informant)	.696	.680	.646	<b>.725</b>	.381	<b>.797</b>	<b>.731</b>
India (Informant)	.478	.479	.488	.629	.088	<b>.730</b>	.631

*Notes.* International informant data does not include the full Big Six scale (only 5 of the 8 items) nor does it include items from HEXACO H, which is why only MAQ scales are listed here. In bold are coefficients that are greater than or equal to .7.

Table 5. Mean of inter-item correlations for informant-report MAQ scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat
U.S. (Student/Informant)	.327	.368	.391	.432	.129	.473	.490
Kenya (Informant)	.209	.116	.225	.379	.052	.360	.290
Sing/HK (Informant)	.365	.347	.314	.397	.134	.495	.404
India (Informant)	.186	.187	.192	.298	.024	.403	.300

Table 6. Variance of inter-item correlations for informant-report MAQ scales in each country

	Compas	Bene	NonM	Fair	Truth	Prom	Grat
U.S. (Student/Informant)	<b>.007</b>	<b>.007</b>	<b>.001</b>	<b>.008</b>	.010	.011	<b>.005</b>
Kenya (Informant)	.035	.015	<b>.003</b>	.025	.033	.030	.019
Sing/HK (Informant)	<b>.003</b>	<b>.003</b>	<b>.005</b>	.028	<b>.004</b>	.011	.021
India (Informant)	.011	<b>.006</b>	.012	.031	.013	<b>.009</b>	.016

*Notes.* In bold are coefficients that are less than .010. For the U.S. informant sample, N ranged from 1022 for the Compassion scale to 1371 for the Truth-Telling measure (based on which items were available from each term of the General Survey).

**Test-Retest Stability.** Test-retest stability of the MAQ was comparable to that of Honesty/Humility of the HEXACO inventory (see Table 7). Not surprisingly, the aggregate scales generally had better cross-time stability than the subscales composed as little as 2 items in the case of some of the HEXACO scales and as many as 4 items for each MAQ scale. The Pearson correlations between Wave 1 and Wave 2 MAQ subscales ranged from .63 (for Non-Maleficence) to .75 for Beneficence, and even the scales with low internal

consistency showed good test-retest stability. The HEXACO facets' test-retest stability ranged from .50 for Sincerity to .81 for Fairness.

Table 7. One-year retest stability for MAQ and HEXACO-H in National Sample

	<i>r</i>	95% CI	df
MAQ Beneficence	.75	[.71, .79]	494
MAQ Non-Maleficence	.63	[.57, .68]	494
MAQ Fairness	.71	[.66, .75]	493
MAQ Truth-Telling	.67	[.62, .71]	494
MAQ Promise-Keeping	.68	[.63, .72]	494
MAQ Gratitude	.64	[.59, .69]	494
MAQ Compassion	.68	[.63, .73]	494
MAQ Tolerance	.67	[.62, .72]	494
MAQ Total	.81	[.78, .84]	493
HEX Fairness	.81	[.74, .87]	109
HEX Sincerity	.50	[.35, .63]	109
HEX Greed Avoidance	.70	[.59, .79]	109
HEX Modesty	.64	[.52, .74]	109
HEXACO H	.83	[.76, .88]	109

**Item Response Theory (IRT).** The main goal of the IRT analysis was to examine the distribution of person estimates on the latent moral trait as estimated using the full MAQ inventory, and as estimated using the H scale from the Big Six. One question was the extent to which the rank-ordering of people would be consistent across the two measures. In addition, the distribution of person estimates was examined for each by country to see whether there were systematic differences in the distributions of the latent trait (as estimated by the MAQ inventory or the H scale from the Big Six) across countries.

Since the MAQ was not sufficiently invariant across countries (see Structural Validity results below), in the aggregated International Sample, parameters in the IRT Rasch family PCM were estimated separately for each country. The model was anchored using one MAQ item considered to be invariant, which was chosen based on its relatively small differences in means across 33

different countries found in Saucier et al. (2015): “I would never take things that aren't mine.”

For the whole 28-item MAQ inventory, overall EAP (expected a priori) reliability in self-reported data was .864 (weighted likelihood estimate or WLE reliability was .858). For the H inventory, overall EAP reliability was .587 in self-reported data (WLE reliability was .541). For the informant-rated MAQ inventory, the overall EAP reliability was .913 (WLE reliability .911).

The distribution of scores across countries was comparable in self-reported data for the latent trait when estimated using the MAQ inventory and the latent trait when estimated using the Big Six H scale (see Figures 1, 2, and 5). However, the distribution as estimated using the MAQ did show slightly more peakedness in Kenya in self-report data, and in both Kenya and India in informant-report data, compared to the other countries, whereas the latent trait estimated using the Big Six-H self-report was distributed more similarly across countries. In addition, though the distributions of the theta estimates for the three measures look fairly similar, they do not match perfectly in their rank-ordering of people on the latent trait— the correlation between the rank-orderings is  $r=.66$ , 95% CI [.63, .69],  $t(1254)=31.28$ ,  $p<.05$ . This lack of agreement could be due in part to the higher person-error for the H dimension, and also due to the fact that they are measuring similar but not completely overlapping traits.

Another useful way to look at how well the measure is doing across all levels of the latent trait is to look at the relationship between the SE and Theta. It appears that both inventories result in more measurement error for higher levels of the latent trait (see Figures 3 and 4). This is also true for the informant-ratings (see Figure 6).

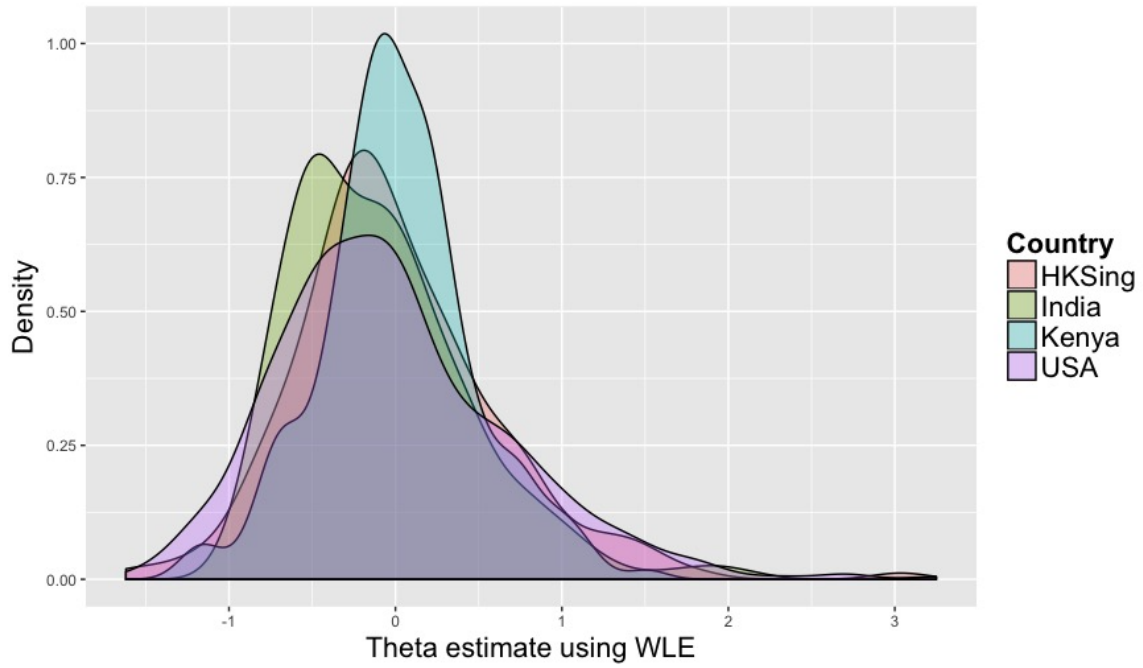


Figure 1. Distribution of Theta estimates using MAQ inventory in self-report data by country

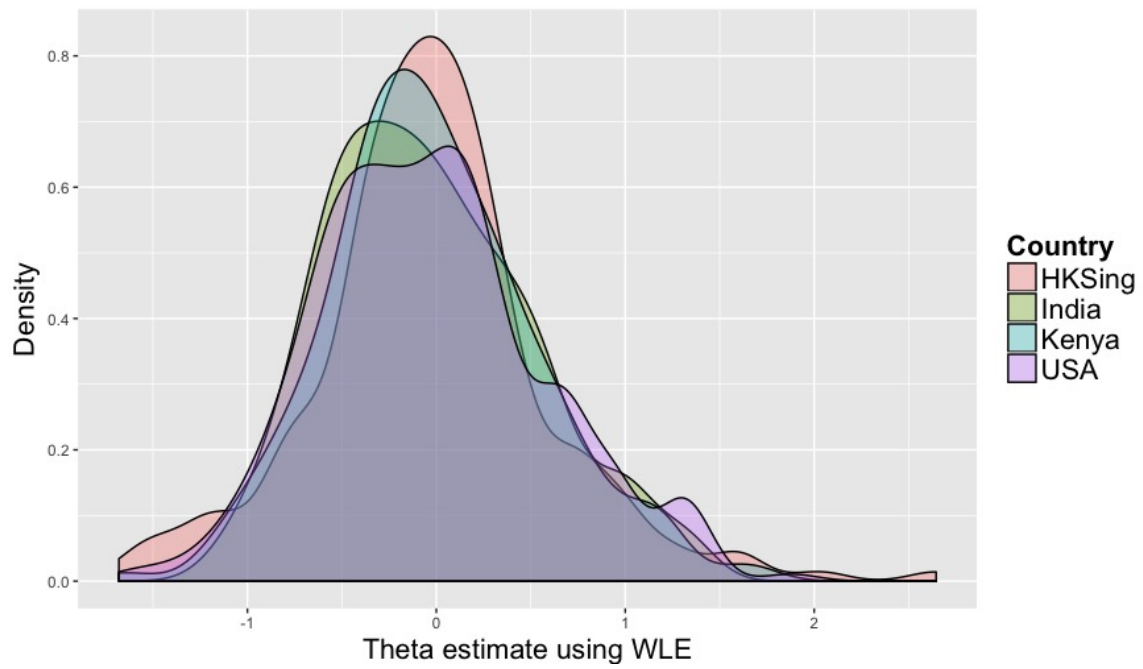


Figure 2. Distribution of Theta estimates using 8-item Big Six H in self-report data by country

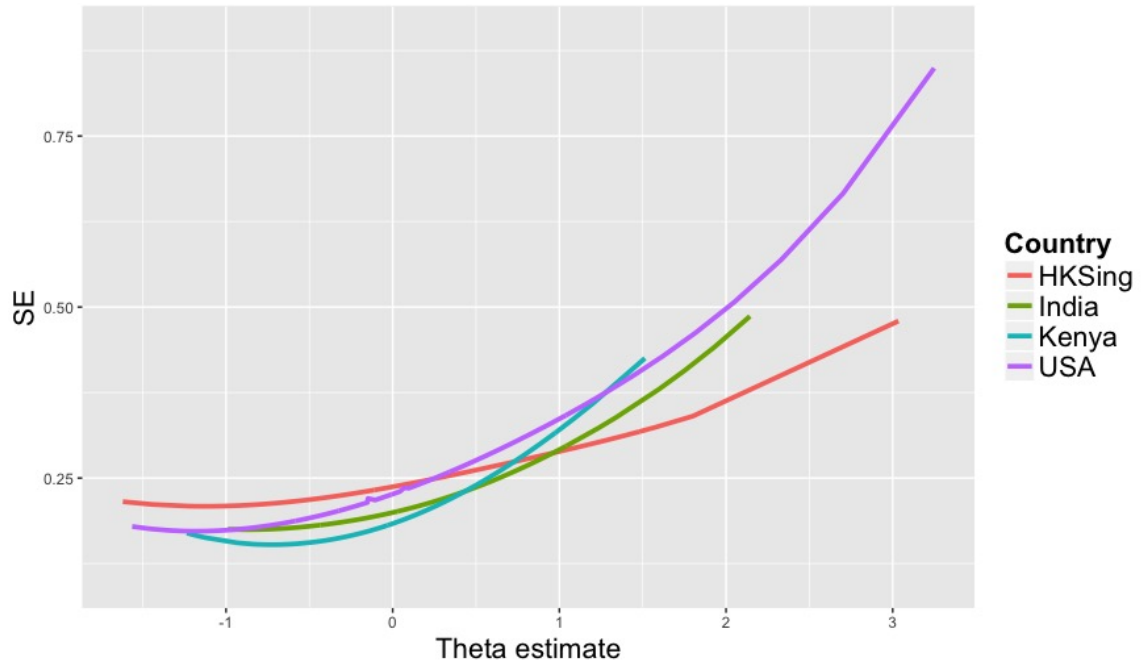


Figure 3. Relationship between Theta and SE for MAQ inventory in self-report data

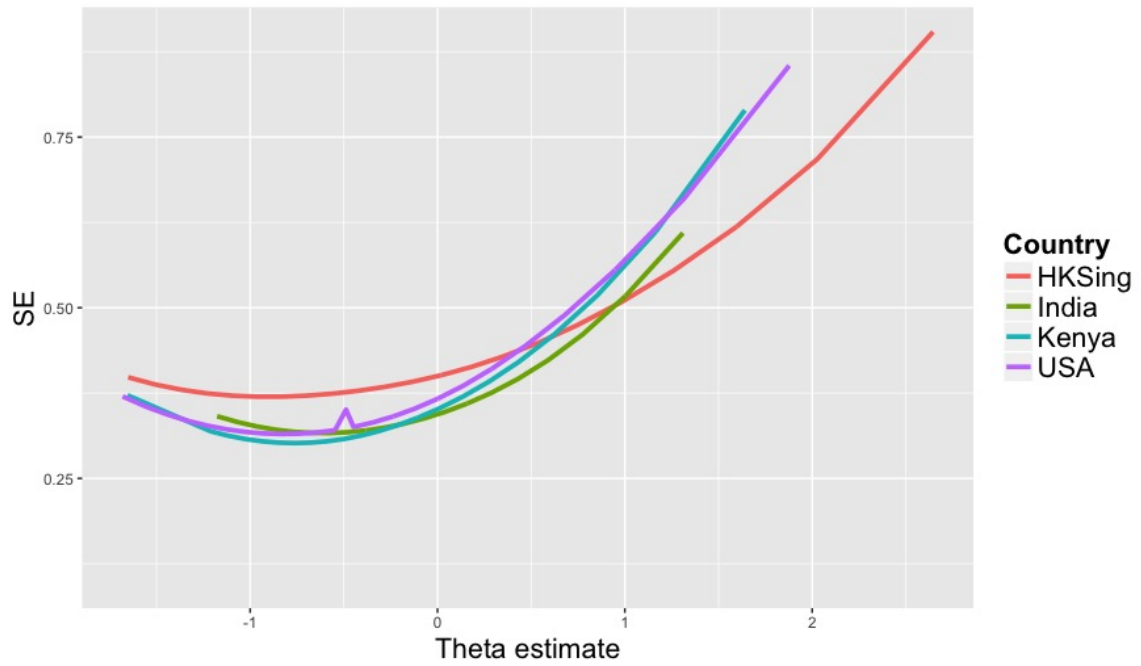


Figure 4. Relationship between Theta and SE for Big Six H in self-report data

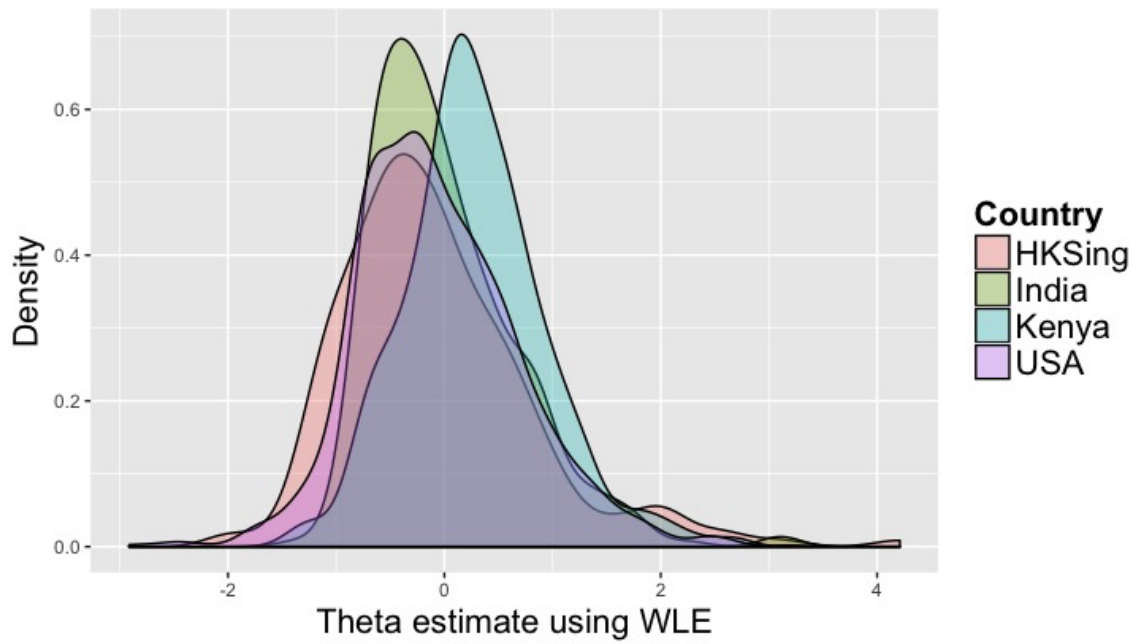


Figure 5. Distribution of Theta estimates using MAQ inventory in informant data by country

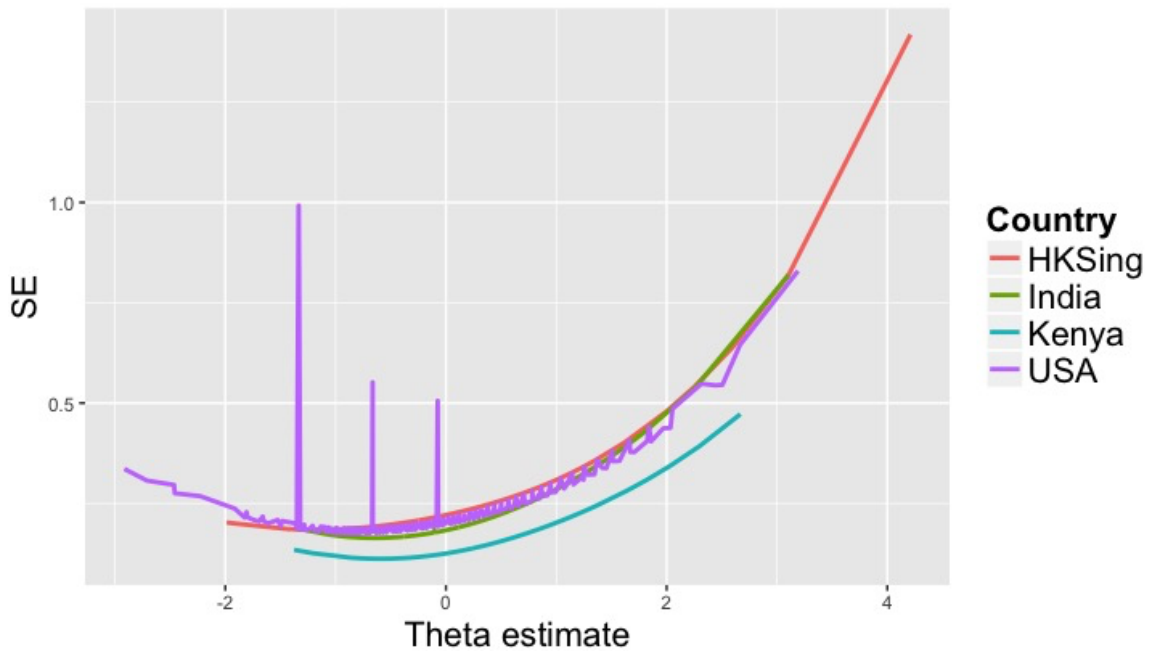


Figure 6. Relationship between Theta and SE for informant-rated MAQ inventory



## Structural Validity

**Structure in Self-Report Data.** A seven-factor model corresponding to each of the seven MAQ domains was tested in the Community Sample. Fit was good for the seven-factor model (see Table 9). However, the factor correlations were used to inform a simpler structure that could be tested in the international samples, due to the potential for model estimation problems that can arise when factors are highly correlated. The pattern of factor correlations (see Table 8) roughly corresponded to a three-factor structure based on the distinction between hortatory/supererogatory ethical duties (praised if done, but no strict ethical obligation to perform: Beneficence, Compassion), minatory ethical duties (ethical duties of strict obligation: Fairness, Non-Maleficence, Truth-Telling, Promise-Keeping), and quasi-moral duties (duties that are conditional on the behavior of others: Gratitude/Reciprocity). However, if one relaxes the assumption that the factor structure should conform to a simple structure, the pattern of correlations and item-level EFA in the training set (U.S. Community Sample) more closely resembled a model where Non-Maleficence and Gratitude/Reciprocity are interstitial to the broader factor centering on hortatory duties (could also be interpreted as a factor centering on care), and the broader factor centering on minatory duties (could also be interpreted as a factor centering on justice, or deontological ethical principles). For this reason, a four-factor model was tested in which Gratitude/Reciprocity and Non-Maleficence were kept separate from these two broader factors. An “Interstitial” model was also tested, where the items were specified to load on two broad factors corresponding to Care and Justice. For this model, Compassion and Beneficence items loaded only on Care, Promise-Keeping, Truth-Telling, and Fairness items were allowed to load only on Justice,

and Gratitude/Reciprocity and Non-Maleficence items were allowed to load on both these factors (i.e., as interstitial between the factors).

In addition, since the moral domain has been modeled in other personality inventories as a single factor and most factor correlations were relatively high, a one-factor model was tested in the international samples.

In the Community Sample, the seven-factor model had the best fit of the models; fit was generally better the more factors the model had (see Table 9). However, the two-factor interstitial model had similar fit to the three-factor models.

Table 8. Estimated factor correlation matrix for seven-factor model in U.S. (Community Sample) data

	Compas	Bene	NonMal	Grat	Truth	Fair
Compas	-	-	-	-	-	-
Bene	.783	-	-	-	-	-
NonMal	.763	.788	-	-	-	-
Grat	.377	.658	.617	-	-	-
Truth	.360	.393	.638	.421	-	-
Fair	.490	.598	.820	.627	.850	-
Promise	.226	.342	.588	.573	.543	.719

Table 9. Fit of seven-, three-, and one-factor models in Community Sample

# factors	Chi-square	df	CFI	AIC	BIC	RMSEA	SRMR
Seven	757.72	329	.846	32731.38	32813.62	.05	.055
Four	906.245	344	.798	32849.904	33112.922	.056	.061
Three*	950.181	347	.784	32887.841	33138.132	.058	.064
Three**	982.345	347	.772	32920.005	33170.296	.06	.065
One	1292.063	350	.662	33223.723	33461.287	.072	.072
Interstitial***	1002.0126	341	.763	32951.686	33227.43	.061	.061

*Notes.* \*For this three-factor model, Non-Maleficence was grouped with Compassion and Beneficence (consistent with the distinction between moral duties centering on not harming/benefiting individual others and moral duties centering on following moral principles—this is the model that was most consistent with results from the item-level EFA). \*\*For the second three-factor model, Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling (consistent with distinction between minatory and hortatory duties). \*\*\*For the interstitial model, items were specified to load on two broad factors corresponding to Care and Justice: Compassion and Beneficence items loaded on one factor, Promise-Keeping, Truth-Telling, and Fairness items loaded on the other factor, and Gratitude/Reciprocity and Non-Maleficence items were allowed to load on both these factors.

Next these models were cross-validated in the international samples and the U.S. National Sample. A seven-factor model was tested in the Hong Kong/Singapore Sample, the Kenya Sample, the India Sample, and the National Sample. Latent variables were scaled by setting their factor variances to 1. In all four samples, the covariance matrix of latent variables was not positive definite. For the four-factor model, only Kenya had a latent variable covariance matrix that was not positive definite (NPD). Both three-factor models successfully ran in all countries (without NPD matrices), and the one-factor and interstitial models also successfully ran in all countries.

For the three-factor model, fit was poorest in India (CFI=.577), followed by Hong Kong/Singapore (CFI=.613), Kenya (CFI=.738), and finally the National Sample (CFI= .758). Fit did not decrease enough from the three-factor to the one-factor model to warrant retaining three factors in Kenya and India (see Table 10); however, fit did worsen considerably for all indices in the Hong Kong/Singapore Sample for the one-factor model relative to the three-factor model.

If one were to choose a three-factor model, which would make the most sense to choose? In other words, is there more empirical support for a model differentiating minatory and hortatory duties, or one differentiating care and justice? Across countries, the three-factor model corresponding to the latter distinction had a slight advantage compared to the three-factor model corresponding to the prior distinction, but this difference is fairly small. It is also possible that the Non-Maleficence scale ended up measuring manners or etiquette, which would more clearly fit with the more care-based moral duties than the minatory duties (one of the four items is “Is sometimes rude to others”).

The interstitial model only did better than the three-factor models in India, and not much better even there. In general, though the scale correlations suggested a rough interstitial structure, it appears that there is no advantage to modeling these factors interstitially at the item-level. This was also true in the Community Sample, where these models were all developed.

If one were interested in defining country-specific models for ethical virtues rather than a model that generalizes across countries, the current data suggested that a model differentiating multiple dimensions of ethical virtue may fit best in Hong Kong/Singapore and the U.S., but did not have much advantage over a simple one-factor model in India and Kenya. In Kenya, fit indices were all about the same for each model, suggesting no structural benefit of modeling ethical virtue with more than one factor; in India, the interstitial model fit slightly better according to CFI, RMSEA and AIC, but was about equal to the other models according to SRMR and BIC. In contrast, in the four-factor model was better by all fit indices in the U.S. National Sample, and the four-factor model was best by CFI in Hong Kong/Singapore (other models with multiple dimensions fit about as well in Hong Kong/Singapore, but the worst model was clearly the one-factor model, suggesting a structural advantage to at least some differentiation of the domains). Though the pattern of correlations did not warrant modeling the MAQ as multiple dimensions in Kenya and India, more dimensions may still aid in prediction in these countries; future research could examine whether the increased differentiation provided by a model with more than one dimension has a predictive advantage, even if there is no benefit in terms of model fit.

Table 10. Fit indices for models in self-report data across countries

Country	Model	Chi-square	df	CFI	AIC	BIC	RMSEA	SRMR
Kenya	Four	NPD	-	-	-	-	-	-
	Three*	693.68	321	.738	16125.240	16314.652	.075	.086
	Three**	692.586	321	.739	16124.146	16313.558	.075	.086
	One	705.329	324	.732	16130.889	16310.332	.076	.086
	Interstitial	733.955	341	.734	16697.560	16913.238	.075	.085
India	Four	1202.747	318	.580	23707.489	23932.069	.094	.114
	Three*	1211.201	321	.577	23709.943	23923.294	.094	.113
	Three**	1215.976	321	.575	23714.718	23928.069	.095	.113
	One	1240.844	324	.564	23733.586	23935.708	.095	.113
	Interstitial	1214.963	341	.598	24567.824	24811.119	.091	.114
HK/Sing	Four	796.281	318	.637	13296.558	13495.351	.086	.088
	Three*	829.679	321	.613	13323.956	13512.809	.088	.091
	Three**	807.679	321	.630	13301.956	13490.809	.086	.088
	One	938.752	324	.533	13427.030	13605.943	.097	.095
	Interstitial	880.881	341	.598	13835.338	14050.696	.088	.089
U.S. (Nat)	Four	1202.601	318	.789	36671.434	36928.146	.072	.071
	Three*	1339.692	321	.758	36802.525	37046.401	.077	.078
	Three**	1320.238	321	.762	36783.071	37026.947	.076	.077
	One	1904.341	324	.624	37361.174	37592.214	.096	.089
	Interstitial	1340.767	316	.756	36813.600	37078.869	.078	.072

*Notes.* No fit indices are listed for the four-factor model in Kenya because there, the latent variable covariance matrix of this model was not positive definite. Three\* factor model is the version where Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling. Three\*\* factor model is the version where Non-Maleficence was grouped with Compassion and Beneficence.

*Measurement Invariance Tests Across Hong Kong/Singapore, India, Kenya, and National Sample.* For the models that successfully ran across all countries, measurement invariance tests were conducted to determine to what extent scale scores can be used to make inferences about relationships between MAQ virtue scales and other constructs across countries.

For all models, there was a moderate drop in all fit indices except for BIC going from the configural model to the factorial model, but an even larger drop from the factorial model to the scalar model (see Tables 11, 12, 13, and 14); this means that the items cannot be assumed to relate to the factors in the same way across countries, and further, item means are not comparable across countries. Of note, the BIC index did suggest the best model was the factorial model, but this was the only fit index indicating this. Like AIC, BIC penalizes model complexity, but does so to a larger degree than AIC. Since all other fit indices (including RMSEA, which also includes a penalty for more complex models) indicated that the factorial model fit worse than the configural model, and the configural model itself did not meet standard thresholds for acceptable fit, it is best to assume no level of invariance holds.

Table 11. Measurement invariance tests for three-factor model (version where Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling) in self-report data

	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	4074.251	1284	.691	90177.664	91902.451	.083	.087
Factorial	4415.712	1365	.663	90357.125	91666.115	.084	.110
Scalar*	5703.790	1437	.528	91501.203	92440.596	.097	.120

\*For scalar test, model in India had a not positive definite (NPD) latent variable covariance matrix

Table 12. Measurement invariance tests for three-factor model (version where Non-Maleficence was grouped with Compassion and Beneficence) in self-report data

	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	4036.479	1284	.696	90139.892	91864.680	.083	.086
Factorial	4387.882	1365	.666	90329.295	91638.285	.084	.109
Scalar*	5677.108	1437	.531	91474.521	92413.914	.097	.119

\*For scalar test, model in India was NPD

Table 13. Measurement invariance tests for one-factor model in self-report data

	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	4789.266	1296	.614	90868.679	92531.867	.093	.092
Factorial	5158.237	1377	.582	91075.650	92323.041	.094	.115
Scalar	6446.980	1455	.448	92208.393	93055.386	.105	.127

Table 14. Measurement invariance tests for interstitial model in self-report data

	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	4011.056	1264	.696	90154.469	91981.922	.083	.085
Factorial	4458.007	1366	.658	90397.420	91701.277	.085	.109
Scalar	5743.089	1441	.524	91532.502	92451.362	.098	.120

**Structure in Informant-Report Data.** An item-level EFA in the Student/Informant Sample was used to develop feasible models that could then be tested in the international informant data (see Appendix B for loadings from the EFA). The EFA suggested a similar pattern for the informant data as it did in the self-report data—where there were broad dimensions characterized by hortatory and minatory duties (which could also be interpreted as care and justice), and Non-Maleficence and Gratitude/Reciprocity were interstitial to these dimensions. In the Student/Informant Sample, the factors were even more highly correlated than in the self-report data (see Table 15 for the factor correlation matrix). In addition to the models tested in the self-report data, a simpler two-factor model was tested in the informant data in which the Gratitude/Reciprocity dimension and items were removed entirely, inasmuch as the items for this dimension loaded about equally on both the broader dimensions. In the Student/Informant Sample, the four-factor model performed the best but fit for all models was similar in this sample (see Table 16).

The four-factor model had a not positive definite (NPD) latent variable covariance matrix in India and Kenya, and both three-factor models had this same issue in India. High correlations were estimated between all three factors in the three-factor model across countries (see Table 18). The version of the three-factor model where Non-Maleficence was grouped with Compassion and Beneficence also had a NPD latent variable covariance matrix in Kenya.

The only models that successfully ran in all countries in the informant data were the two-factor model (without Gratitude/Reciprocity) and the one-factor model. The one-factor model fit similarly or better than the two-factor model in India and Kenya, but the two-factor model clearly fit better in Hong



Kong/Singapore (and in the Student/Informant Sample). See Table 17 for fit indices for each model in each country. Measurement invariance tests showed a similar pattern as they did in self-report data, suggesting that factorial invariance and scalar invariance cannot be assumed (see Tables 19 and 20). As in the self-report data, BIC indicated the factorial model had better fit than the configural model, but again, since even configural invariance could not be assumed and the other fit indices suggested there was a decrement in fit moving from configural to factorial, measurement invariance on any level should not be assumed.

Table 15. Estimated factor correlation matrix in Student/Informant Sample (where latent variable covariance matrix was not positive definite)

	Compas	Bene	NonMal	Grat	Truth	Fair
Compas	-	-	-	-	-	-
Bene	.984	-	-	-	-	-
NonMal	.944	.891	-	-	-	-
Grat	.772	.946	.782	-	-	-
Truth	.751	.643	.813	.754	-	-
Fair	.840	.865	.888	.918	.938	-
Promise	.578	.758	.681	.828	.682	.860

Table 16. Fit of models in Student/Informant Sample

# factors	Chi-square	df	CFI	AIC	BIC	RMSEA	SRMR
Seven (NPD)	-	-	-	-	-	-	-
Four	1906.787	344	.831	62288.862	62582.48	.073	.059
Three*	1953.564	347	.826	62329.639	62609.05	.074	.060
Three**	2079.537	347	.813	62455.612	62735.024	.077	.062
Two****	1645.915	251	.822	60836.833	61074.281	.077	.062
One	2282.652	350	.791	62652.727	62917.931	.081	.063
Interstitial***	1930.088	341	.828	62318.163	62625.989	.074	.059

*Notes.* \*For the first three-factor model, Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling (consistent with the distinction between minatory and hortatory duties). \*\*For the second three-factor model, Non-Maleficence was grouped with Compassion and Beneficence, consistent with the distinction between moral duties centering on not harming/benefiting individual others (care) and moral duties centering on following moral principles (justice). \*\*\*For the interstitial model, items were specified to load on two broad factors corresponding to care and justice; Compassion and Beneficence items loaded on one factor, Promise-Keeping, Truth-Telling, and Fairness items loaded on the other factor, and Gratitude/Reciprocity and Non-Maleficence items were allowed to load on both these factors.\*\*\*\*The two-factor model did not include the Gratitude/Reciprocity factor or items, and had two broad factors that could be interpreted as care and justice.

Table 17. Fit indices for models in informant-report data across countries

	Model	Chi-square	df	CFI	AIC	BIC	RMSEA	SRMR
India	Four	NPD	-	-	-	-	-	-
	Three*	NPD	-	-	-	-	-	-
	Three**	NPD	-	-	-	-	-	-
	Two	943.831	251	.713	21573.642	21757.049	.094	.099
	One	1253.571	350	.706	24843.735	25053.164	.091	.097
	Interstitial	1144.879	341	.739	24753.044	24996.13	.087	.096
HK/Sing	Four	823.427	344	.789	14458.758	14663.563	.083	.074
	Three*	835.474	347	.785	14464.806	14659.701	.084	.074
	Three**	855.155	347	.776	14484.486	14679.381	.085	.076
	Two	588.648	251	.811	12596.853	12758.715	.082	.071
	One	948.590	350	.736	14571.922	14756.907	.092	.076
	Interstitial	821.950	341	.788	14463.282	14677.996	.084	.073
Kenya	Four	NPD	-	-	-	-	-	-
	Three*	NPD	-	-	-	-	-	-
	Three**	810.237	347	.734	16600.777	16795.082	.082	.087
	Two	640.263	251	.721	14420.477	14581.849	.088	.091
	One	819.162	350	.731	16603.701	16788.126	.082	.087
	Interstitial	-	-	-	-	-	-	-

*Notes.* In Kenya, the standard errors of the interstitial model could not be computed. \*For the first three-factor model, Non-Maleficence was grouped with Fairness, Promise-Keeping, and Truth-Telling (consistent with the distinction between minatory and hortatory duties). \*\*For the second three-factor model, Non-Maleficence was grouped with Compassion and Beneficence, consistent with the distinction between moral duties centering on not harming/benefiting individual others (care) and moral duties centering on following moral principles (justice).

Table 18. Estimated factor correlation matrix for three-factor model in each country

	Hort-Mina	Hort-Grat	Mina-Grat
U.S. (Student/Informant)	.864	.872	.908
India (informant)	.850	.986	.995
HK/Sing (informant)	.823	.902	.796
Kenya (informant)	.909	.968	.942

*Note.* Hort-Mina refers to the correlation between Hortatory and Minatory factors, Hort-Grat refers to the correlation between Hortatory and Gratitude/Reciprocity factors, and Mina-Grat refers to the correlation between Minatory and Gratitude/Reciprocity factors.

Table 19. Measurement invariance tests for two-factor model in informant-report data

	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	3818.657	1004	.791	109619.806	111199.451	.082	.071
Factorial	4261.161	1076	.763	109918.310	111108.453	.085	.091
Scalar	5429.547	1142	.681	110954.696	111298.563	.095	.101

Table 20. Measurement invariance tests for one-factor model in informant-report data

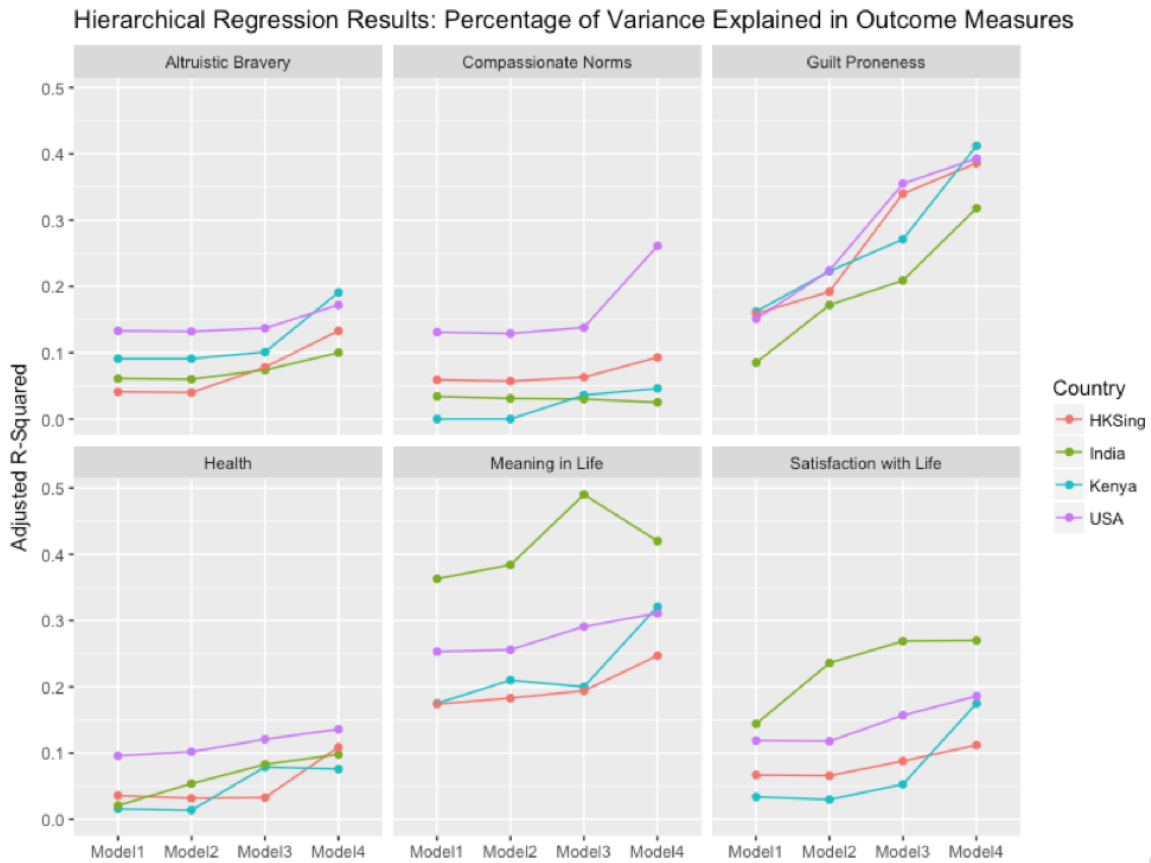
	Chi-square	Df	CFI	AIC	BIC	RMSEA	SRMR
Configural	5303.975	1400	.761	118896.085	120692.994	.085	.072
Factorial	5763.134	1484	.738	119187.245	120534.926	.086	.093
Scalar	7130.246	1565	.659	120392.356	121306.855	.096	.103

## Predictive Validity

The main research question this section addresses is how various candidate measures for ethical virtue (Agreeableness and Conscientiousness from the Big Five, the appended Honesty dimension to the Big Five, the four HEXACO facets, and the seven MAQ subscales) compared in their prediction of outcomes— are they mainly redundant with one another, or are some of them stronger predictors where others fall short? Across all outcomes, which ethical virtue measures are the strongest predictors?

To answer this main research question, hierarchical regression analyses were used, entering measures for ethical virtue in successive order for each outcome (see Figure 7). Model 1 predicted the outcomes with only Agreeableness and Conscientiousness scales from the Big Five; if there were an increase from 0 for Model 1 and a flat line after, this would suggest the ethical virtue model gleaned from the Big Five is the best we have for predicting for these outcomes. This would imply the more recent attempts at integrating moral content into models of personality are not useful for predicting these outcomes in these samples. It does appear that Model 1 predicts variation in these outcomes, but the line is not flat after Model 1, suggesting a benefit of not relying solely on the Big Five.

Model 2 adds the Honesty/Propriety scale that was appended to the Big Five. With the exception of guilt proneness, adding the H scale does not generally aid in prediction of outcomes above and beyond Big Five Agreeableness and



*Figure 7.* Adjusted R-squared for hierarchical models predicting each outcome *Notes.* Model 1’s predictors include only Big Five Agreeableness and Conscientiousness. Model 2’s predictors add the H scale appended to the Big Five. Model 3’s predictors add the four H HEXACO facets (Sincerity, Fairness, Greed Avoidance, and Modesty). Model 4’s predictors add the seven MAQ subscales (Beneficence, Non-Maleficence, Compassion, Fairness, Truth-Telling, Promise-Keeping, and Gratitude/Reciprocity).

Conscientiousness. The H scale does aid in predicting health, meaning in life, and satisfaction with life, but only in India. The general lack of added benefit to using the H scale to predict outcomes may be due to the fact that the ethical-virtue validity outcomes chosen for this study are more about helping others than following ethical rules, a core component of Agreeableness rather than Honesty/Propriety.

Model 3 adds the four HEXACO facets, and Model 4 adds the seven MAQ subscales. Increases in adjusted R-squared for Models 3 and 4 may suggest an advantage for facet models (confounded with any strengths of these particular inventories), as these are the only inventories used here which are modeled at the facet level rather than as broad domains. And any increases observed in adjusted R-squared from Model 3 to Model 4 indicate that the MAQ subscales confer an advantage over the HEXACO facets. It does appear as though there is a modest general advantage to modeling ethical virtues using facets. For the HEXACO facets in particular, predictive power is gained in some countries for some outcomes (satisfaction with life and guilt proneness in all countries, health and compassionate norms in Kenya only).

The results for Model 4 suggest the MAQ subscales sometimes confer an additional advantage equal to or greater than that conferred by adding the HEXACO facets. The MAQ aids in prediction of altruistic bravery and guilt proneness across all countries. The MAQ also aids in prediction of compassionate norms in Hong Kong/Singapore and the U.S., health in Hong Kong/Singapore, and meaning in life and satisfaction with life in all countries except India.

In summary, the results from the hierarchical regressions suggest that a fuller measure of ethical virtue, derived from a rationale-driven approach to

identifying ethical virtues, did have an advantage in predicting relevant outcomes. This cannot solely be explained by the advantage conferred by using facets, as the HEXACO inventory was modeled with facets and the MAQ aids in prediction even above HEXACO facets.

Though sheer predictive power is a useful criterion in itself, to more fully evaluate validity, it is also important to determine 1) whether the relationships are in the expected direction, 2) whether the pattern of relationships is as expected, in terms of relative magnitude (discriminant validity), 3) even if the expected relationships are found, support for validity can be strengthened by attempting to rule out alternative explanations for these relationships such as socially desirable responding, and 4) given the study from Westfall and Yarkoni (2016), whether the results showing incremental validity still hold when measurement error is taken into account.

With respect to 1), most correlations were in the direction expected. However, as 2) mentions, to further bolster support for these inventories' validity, it is also important to determine to what extent the relative magnitude of pattern of correlations is what was expected (discriminant validity). One prediction was that the virtues centering on helping others (Beneficence and Compassion) should predict the validity outcomes related to helping others (welfare contribution, compassionate norms, altruistic bravery) more strongly than the rule-focused ethical virtues such as Promise-Keeping and Fairness. These predictions were roughly confirmed in the U.S. National Sample. In the U.S. National Sample, Beneficence was the virtue that was the most strongly related to compassionate norms, welfare contribution and altruistic bravery, while Truth-Telling and Promise-Keeping were the most weakly related. Though not predicted, it is

interesting to note that guilt proneness was the outcome with the most consistent relations across all the ethical virtues.

Similarly, in Hong Kong/Singapore, Beneficence was the virtue most highly related to the validity outcomes focused on helping others, and guilt proneness showed a more consistent relationship with all the ethical virtues. In Kenya and India, there was a different pattern. In Kenya, Fairness, Non-Maleficence, and Gratitude/Reciprocity were most strongly related to welfare contribution, guilt proneness, and altruistic bravery, but Beneficence was most strongly related to compassionate norms. In India, Non-Maleficence and Promise-Keeping were the most strongly related to the outcomes having to do with helping others; guilt proneness again was consistently related to all virtues, but most strongly related to Gratitude/Reciprocity.

With respect to 3), there was some evidence that some correlations were inflated by socially desirable responding, particularly in the international samples. In India, the MAQ aggregate scale did relate more strongly than the HEXACO-H domain scale to the ethical virtue outcomes predicted a priori to be related (welfare contribution, compassionate norms, altruistic bravery, and guilt proneness). Though the MAQ-total score overall predicted outcomes more strongly than HEXACO-H, it is also the case that the correlation between the total MAQ score and the item “is an ethical/moral person” was more highly attenuated for the MAQ when other desirable characteristics are taken into account than for HEXACO-H, suggesting that socially desirable responding may play a role for the MAQ more than for the HEXACO-H. This is not surprising given that the items in the HEXACO are framed as hypothetical situations; responding in a less virtuous way does not actually require one to directly admit that one has a

tendency to violate ethical standards, just that one would be tempted to do so if the opportunity arose.

In Hong Kong/Singapore and Kenya, the MAQ again generally tended to have stronger correlations with the validity outcomes than HEXACO-H, and often comparable correlations with the validity outcomes as the Big Six scales, but both MAQ and the Big Six scales showed larger drops in the correlation with “is an ethical/moral person” once other desirable characteristics were controlled for, while the correlation between HEXACO-H and this item were not affected as much.

In the U.S. National Sample, the zero-order correlations between the aggregate MAQ score and the ethical virtue outcomes were clearly higher than those for the other ethical virtue measures. Further, since the correlation between “is an ethical/moral person” was not attenuated much when other correlations were controlled for, socially desirable responding is a less likely explanation for the MAQ’s superior prediction of validity outcomes. In summary: although the MAQ generally predicts outcomes better than HEXACO, the relation between the MAQ scales and outcomes may at least in part be inflated by socially desirable responding in the international samples, which is less the case in the U.S. National Sample.

With respect to 4), the hierarchical regressions were also run using SEM so that measurement error could be incorporated into the models. Westfall and Yarkoni (2016) found that incremental validity claims often do not hold up once measurement error is taken into account, and counter-intuitively, this is particularly true when reliability is moderate rather than very low or very high. Since it was not possible to model these highly correlated scales as latent variables without arriving at NPD latent variable covariance matrices, following Bollen



(2009), scale scores were used as indicators for each scale, and measurement error was explicitly incorporated into the model by setting the variance of the error term corresponding to the latent variable underlying that scale equal to  $(1-\alpha)*\text{var}(\text{scale})$ , where  $\alpha$  was the empirical  $\alpha$  estimate for that scale in that particular sample. Unfortunately, these models only ran without estimation problems in the National Sample, and even there only the four models with the multiple-item outcomes (meaning in life, satisfaction with life, guilt proneness, and compassionate norms) converged without estimation problems. For those four outcomes, the percentage of variance explained by the predictors in each outcome looked similar to the results for the models assuming no measurement error (See Figure 8).

For the international samples, participants were asked to rate someone who they admire on the MAQ, so one final validity indicator for the MAQ was whether people who are themselves higher in ethical virtue tended to admire people who are also higher in ethical virtue. There was a strong to moderate correlation between self-reported aggregate MAQ score and informant-rated aggregate MAQ score in all three samples: in India, people with higher aggregate self-reported MAQ scores also tended to admire people who they rated as being higher in MAQ virtue,  $r=.72$ ,  $t(309)=18.06$ ,  $p<.05$ ; in Kenya,  $r=.66$ ,  $t(196)=12.30$ ,  $p<.05$ ; in Hong Kong/Singapore,  $r=.35$ ,  $t(199)=5.24$ ,  $p<.05$ . The finding that people higher in ethical virtue tended to admire people who were also higher in ethical virtue is consistent with past research by Schlenker, Weigold, and Schlenker (2008) which found that people higher in integrity tended to have heroes with moral qualities (principled, benevolent, and spiritual) while the heroes of people who were lower in integrity tended to have heroes with admirable qualities unrelated to morality.

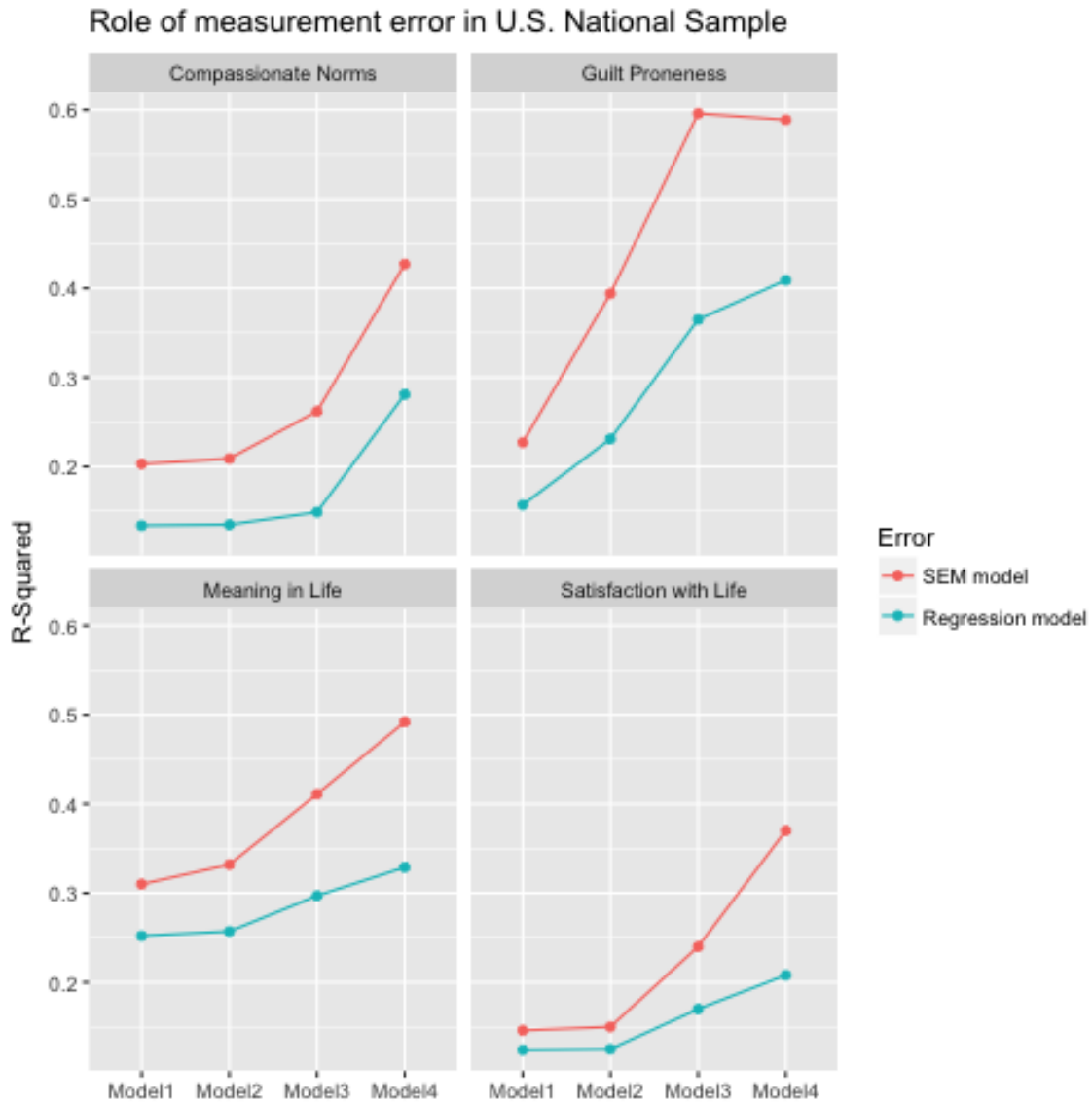


Figure 8. R-squared estimates for the regression model assuming no measurement error and SEM regression model.

Notes. Model 1's predictors include only Big Five Agreeableness and Conscientiousness. Model 2's predictors add the H scale appended to the Big Five. Model 3's predictors add the four H HEXACO facets (Sincerity, Fairness, Greed Avoidance, and Modesty). Model 4's predictors add the seven MAQ subscales (Beneficence, Non-Maleficence, Compassion, Fairness, Truth-Telling, Promise-Keeping, and Gratitude/Reciprocity). For the ethical-virtue scales, the amount of variance due to measurement error was explicitly added to the model and was calculated based on each scale's alpha estimate in the U.S. National Sample. All outcomes were modeled as latent variables.

Table 21. Zero-order correlations between ethical virtue scales and outcomes in U.S. National Sample

	Welfare Contrib	Comp Norms	Guilt Prone	Shame Prone	Altruist Bravery	MIL	SWL	Health	Ident Ethical	Ident Ethical (- Desir)
Big Six-H	.13	.20*	.44*	.16*	.21*	.22*	.17*	.09	.44*	.40*
Big Six-A	.36*	.36*	.37*	.05	.34*	.45*	.30*	.22*	.37*	.33*
Big Six-C	.22*	.20*	.25*	-.10	.26*	.36*	.28*	.28*	.35*	.27*
HEX- H (domain)	.15*	.18*	.45*	.09	.20*	.16*	.11	.03	.31*	.30*
HEX- Sincerity	.09	.13	.25*	.03	.14	.02	.02	-.03	.13	.13
HEX- Fairness	.24*	.25*	.55*	.13	.21*	.29*	.20*	.11	.41*	.36*
HEX- Greed	.08	.08	.09	-.06	.03	.09	.13	.00	.09	.11
HEX- Modesty	.01	.11	.18*	.04	.04	-.06	-.09	-.13	.09	.12
MAQ- Total	.40*	.44*	.54*	.11	.36*	.41*	.23*	.16*	.55*	.51*
MAQ- Comp	.34*	.37*	.33*	.10	.25*	.20*	.05	-.02	.26*	.24*
MAQ- Bene	.43*	.50*	.37*	.12	.36*	.37*	.18*	.10	.35*	.28*
MAQ- NonMal	.34*	.31*	.47*	.14	.30*	.37*	.23*	.21*	.44*	.38*
MAQ- Grat	.27*	.34*	.34*	.03	.25*	.24*	.10	.09	.37*	.29*
MAQ- Truth	.10	.13	.33*	.06	.10	.14	.08	.04	.37*	.33*
MAQ- Fair	.21*	.24*	.43*	.09	.23*	.28*	.16*	.13	.43*	.36*
MAQ- Promise	.19*	.23*	.32*	-.04	.24*	.36*	.29*	.24*	.45*	.37*

Note. Pairwise deletion used, N=528-536. \*p<.001 (for 526 df, all Pearson R values greater than |.14| meet this threshold for significance)

Table 22. Zero-order correlations between ethical virtue scales and outcomes in India Sample

	Welfare Contrib	Comp Norms	Guilt Prone	Shame Prone	Altruist Bravery	MIL	SWL	Health	Ident Ethical	Ident Ethical (- Desir)
Big Six-H	.06	.11	.42*	-.09	.18	.24*	-.06	-.06	.35*	.21*
Big Six-A	.26*	.13	.29*	-.18	.24*	.56*	.25*	.12	.44*	.20*
Big Six-C	.34*	.20*	.26*	-.06	.24*	.53*	.38*	.16	.49*	.22*
HEX- H (domain)	.10	.14	.39*	-.09	.18	.29*	-.07	-.07	.36*	.29*
HEX- Sincerity	.10	.06	.29*	-.01	.14	.19*	.02	.02	.25*	.14
HEX- Fairness	.15	.16	.38*	-.07	.21*	.36*	.03	-.07	.43*	.33*
HEX- Greed	.17	.04	.03	-.21*	.07	.16	.24*	.01	.08	.06
HEX- Modesty	-.14	-.02	.08	-.20*	-.09	-.07	-.18	-.23*	-.09	.07
MAQ- Total	.17	.15	.54*	-.13	.28*	.47*	.10	.01	.47*	.28*
MAQ- Comp	.03	.05	.37*	-.11	.07	.28*	.01	-.08	.25*	.14
MAQ- Bene	.10	.16*	.39*	-.14	.21*	.31*	.03	-.02	.29*	.15
MAQ- NonMal	.25*	.13	.38*	-.06	.24*	.42*	.22*	.12	.29*	.13
MAQ- Grat	.08	.13	.43*	-.09	.21*	.37*	.02	.05	.41*	.24*
MAQ- Truth	-.04	-.04	.31*	-.04	.04	-.03	-.11	-.19*	-.03	.01
MAQ- Fair	.15	.13	.37*	-.11	.29*	.39*	.07	.04	.43*	.24*
MAQ- Promise	.21*	.16	.36*	-.09	.27*	.48*	.21*	.08	.57*	.35*

Note. Pairwise deletion used, N=310-312. \*p<.001 (for 308 df, all Pearson R values greater than |.19| meet this threshold for significance)

Table 23. Zero-order correlations between ethical virtue scales and outcomes in Hong Kong/Singapore Sample

	Welfare Contrib	Comp Norms	Guilt Prone	Shame Prone	Altruist Bravery	MIL	SWL	Health	Ident Ethical	Ident Ethical (- Desir)
Big Six-H	.11	.06	.32*	.15	.06	.11	.13	.13	.28*	.23
Big Six-A	.36*	.27*	.36*	.06	.21	.35*	.25*	.01	.27*	.14
Big Six-C	.19	.08	.16	-.02	.10	.35*	.27*	.18	.25*	.11
HEX- H (domain)	.23	.13	.46*	.14	.17	.21	.12	.01	.18	.15
HEX- Sincerity	.09	.07	.32*	.07	.20	.13	.07	.05	.17	.14
HEX- Fairness	.29*	.23	.50*	.19	.19	.30*	.23*	.08	.23	.13
HEX- Greed	.13	.19	.24*	-.01	.23*	.18	.25*	.15	.04	.01
HEX- Modesty	-.01	.12	.29*	-.08	.11	.09	-.02	-.05	-.04	.03
MAQ- Total	.31*	.26*	.52*	.12	.31*	.42*	.22	.13	.45*	.36*
MAQ- Comp	.26*	.22	.27*	.13	.21	.22	.09	.09	.08	.04
MAQ- Bene	.36*	.35*	.40*	.00	.39*	.32*	.12	.04	.26*	.14
MAQ- NonMal	.21	.15	.35*	.13	.17	.30*	.08	-.05	.32*	.21
MAQ- Grat	.22	.19	.33*	.03	.24*	.33*	.14	.16	.36*	.27*
MAQ- Truth	-.07	-.03	.31*	.14	.05	.10	.04	.14	.20	.16
MAQ- Fair	.19	.15	.40*	.07	.18	.28*	.23	.18	.39*	.34*
MAQ- Promise	.22	.15	.33*	.09	.15	.40*	.31*	.21	.48*	.41*

Note. Pairwise deletion used, N=310-312. \*p<.001 (for 308 df, all Pearson R values greater than |.19| meet this threshold for significance)

Table 24. Zero-order correlations between ethical virtue scales and outcomes in Kenya Sample

	Welfare Contrib	Comp Norms	Guilt Prone	Shame Prone	Altruist Bravery	MIL	SWL	Health	Ident Ethical	Ident Ethical (- Desir)
Big Six-H	.23	-.01	.47*	.09	.27*	.45*	.17	.07	.44*	.30*
Big Six-A	.30*	.01	.41*	.14	.29*	.42*	.18	.16	.41*	.27*
Big Six-C	.30*	-.01	.31*	.09	.27*	.36*	.22	.14	.49*	.33*
HEX- H (domain)	.17	.07	.33*	.03	.22	.20	.01	.01	.31*	.26*
HEX- Sincerity	.14	.01	.31*	.12	.08	.08	-.02	.05	.14	.13
HEX- Fairness	.16	-.01	.38*	.05	.28*	.30*	.13	-.11	.47*	.38*
HEX- Greed	.18	.05	.17	-.04	.05	.13	.15	.05	.09	.06
HEX- Modesty	.13	.26*	.21	-.17	.12	.12	-.09	-.24*	-.01	.07
MAQ- Total	.36*	.14	.63*	.17	.40*	.52*	.13	-.01	.48*	.34*
MAQ- Comp	.19	.15	.33*	.03	.18	.19	-.03	-.03	.16	.10
MAQ- Bene	.20	.20	.38*	-.01	.16	.19	-.10	-.13	.21	.11
MAQ- NonMal	.33*	.08	.54*	.12	.42*	.46*	.18	.07	.46*	.34*
MAQ- Grat	.27*	.12	.52*	.18	.42*	.49*	.17	-.03	.42*	.23*
MAQ- Truth	.13	-.05	.31*	.12	.15	.27*	-.07	.03	.15	.05
MAQ- Fair	.36*	.07	.57*	.23	.34*	.46*	.22	.03	.47*	.34*
MAQ- Promise	.26*	.05	.44*	.16	.27*	.49*	.30*	.05	.47*	.31*

Note. Pairwise deletion used, N=196-205. \*p<.001 (for 194 df, all Pearson R values greater than |.23| meet this threshold for significance)

**Forecasting.** Forecasting power of the MAQ was examined by looking at the relationship between MAQ scores in Wave 1 of the National Sample and an objective measure of ethical virtue evaluated in Wave 2 of the National Sample which was not used to select items for each domain– Time to return a paid-in-advance survey. I predicted that those who were higher in ethical virtue as measured by the MAQ would be more likely to return the paid-in-advance survey as the sense of dutifulness should compel them to return the survey even though they received no extra reward for doing so. In addition, among people who returned the survey, I expected that ethical virtue should lead people to return the survey in a more timely fashion, controlling for other factors that may potentially affect time to return the survey (interest in the survey and how busy they were).

The MAQ subscales (self-report) did not significantly improve the model predicting whether or not the survey was returned relative to the null model,  $\chi_2(7)=9.17$ ,  $p=.24$ . The aggregated MAQ score (self-report) also did not significantly predict whether the survey was returned and the odds of returning the survey were .998 times that for each unit increase in the aggregate MAQ score. Among people who returned the survey, the strongest predictors of time to return the survey (log-transformed) were interest in the survey and how busy they were (both positively related, meaning that being more busy and more interested tended to relate to taking longer to return the survey), multiple  $R^2=.03$  (adjusted  $R^2=.01$ ),  $F(2, 124)=1.78$ ,  $p=.17$ . The model including those control variables and MAQ virtues had a  $R^2=.04$  (adjusted  $R^2= -.04$ ) and the MAQ virtues did not explain a significant portion of variance above and beyond the control variables,  $F(7, 117)=.16$ ,  $p=.99$ .

One might expect informant-report to be more valid and better predict single behaviors. Using the participants who had informant-ratings on the MAQ (aggregating informant ratings for those who had multiple informants), a logistic regression was run predicting the same outcome. The informant-report MAQ ratings more strongly predicted whether the main participant returned the survey than self-report ratings. Relative to the null model, the informant-rated MAQ scales marginally improved model fit  $\chi_2(7)=10.154$ ,  $p=.18$ , and while the null model correctly classified 73% of cases, this regression with the informant-rated MAQ scales correctly classified 74%. In the multiple-logistic regression, the significant predictors were Beneficence and Compassion; people who were more highly rated in Beneficence tended to be more likely to return the survey (odds ratio=1.3), while people who were higher in Compassion tended to be less likely to return the survey (odds ratio=.81), taking into account the effect of the other variables. Promise-Keeping and Truth-Telling, the two scales one might expect to be most strongly related to this single act of dutifulness, showed no relationship to the likelihood to return the survey, and the overall aggregate informant-rated MAQ measure also did not improve prediction relative to the null model,  $\chi_2(1)=.09$ ,  $p=.77$ .

Among people who returned the survey, the informant-rated MAQ scales as a whole did not significantly predict how long it took to return the survey,  $F(7, 74)=.968$ ,  $p=.46$ ,  $R^2=.08$ , adjusted  $R^2=-.003$ . The informant-rated MAQ scales with the largest coefficients were Non-Maleficence ( $b=-.19$ ) and Compassion ( $b=.25$ ) suggesting that those higher in Compassion took longer to return the survey and those higher in Non-Maleficence took less time to return the survey.



When how busy people were and interest were controlled for, these coefficients stayed about the same.

Since this is just a single behavior rather than a reliable aggregate of behavior after many receipts of paid-in-advance surveys, it is not surprising that the effects found were small. These findings do suggest that the informant-rated MAQ scale (at least the Beneficence, Compassion, and Non-Maleficence subscales) are slightly better predictors of this single behavior than the self-report subscales.

In summary, the main question this dissertation addressed was how different models for ethical virtue compare to one another on a variety of psychometric and validity criteria. Special interest was in the performance of the MAQ, as this inventory is distinct from the others in its derivation from a rationale-based approach to identifying the most fundamental ethical virtues worthy of measurement. Main findings were that the MAQ aids in prediction of outcomes beyond other inventories across countries, with the caveat that this may in part reflect inflation due to socially desirable responding in the international samples. The results also suggest the MAQ scales need improvement before they can be used for cross-cultural measurement, and could also be improved to increase precision for those measuring at the higher levels of ethical virtue. These findings will be elaborated on in the discussion.

## CHAPTER IV

### DISCUSSION

This dissertation contributed in five main ways to the assessment of ethical virtue. First, and perhaps most importantly, this dissertation investigated whether the increased breadth of content represented by the Moral Actions Questionnaire contributed to better prediction of validity criteria and other important outcomes such as life satisfaction, meaning in life, and health, beyond what already-existing personality inventories predict. The Moral Actions Questionnaire did predict some validity outcomes above and beyond what the already-existing inventories predict, but some caution should be exercised in interpreting these results in the international samples due to the evidence found for socially desirable responding (that is, correlations between ethical virtue scales and “is an ethical/moral person” being more attenuated for the MAQ than for the H-HEXACO scale when other desirable attributes were accounted for). In terms of the relationship between ethical virtue and meaning and satisfaction with life, the MAQ scales did aid in prediction above other scales in most samples (i.e., in the U.S. and Hong Kong/Singapore samples, and especially in Kenya, but not in India). Across countries, ethical virtue was consistently positively related to meaning in life, but the relationship with satisfaction with life was weaker and more inconsistent. The single-item self-report indicator of health status tended to be weakly but positively related to ethical virtue across countries.

One explanation for the finding that ethical virtue relates to meaning in life (but not satisfaction with life) is that people who are more virtuous tend to organize their lives around ethical principles. This would be expected to create higher levels of meaning, but not necessarily higher levels of life satisfaction,

since constraining oneself to ethical behavior is sometimes less enjoyable than the alternative. Another explanation for the relationship found between meaning in life and ethical virtue has to do with social relatedness. Multiple studies have found that meaning in life is related to whether social needs such as sense of belonging and need for relatedness have been met (King, Heintzeman, & Ward, 2016). It is possible that ethical virtue is related to having more fulfilling relationships, through its impact on the establishment of mutual trust and reciprocity. If that were the case, people higher on ethical virtue may have more meaningful lives due to their needs for social relatedness being met to a higher degree. However, this explanation does not necessarily account for why ethical virtue would have a stronger relationship with meaning in life than satisfaction with life.

The forecasting analysis suggested that informant responses may be more valid than self-report responses, an interpretation consistent with past research suggesting that for highly evaluative traits, others are more accurate than the self, i.e., Vazire (2010); Vazire and Mehl (2008). More specifically, the informant MAQ scales (but not the self-reported MAQ scales) very slightly contributed to a forecasting prediction of whether a paid-in-advance survey would be returned or not. However, the MAQ scales most predictive of whether or not the survey was returned were not those relating primarily to some moral sensibility related to dutifulness, as was expected. It is possible that the mechanism hypothesized to relate to returning the survey— the sense of dutifulness— was incorrect, and it is actually a sense of reciprocity or beneficence toward the research team that led people to return the survey. And though not originally predicted, it is possible that having a high level of compassion (as perceived by knowledgeable informants) may lead people to spend a lot of time attending to the needs of

others, distractions that could lead them to neglect obligations that seem less pressing.

A second way in which this dissertation contributed to the assessment of ethical virtue has to do with its investigation of the psychometric quality of measures for seven conceptually distinct ethical virtues across self- and informant-report data, across five different countries, across time, and across varying levels of the latent trait. Further, on many of these criteria the MAQ was compared to already-existing personality traits that can be considered competing models of ethical virtue that are already integrated into personality models.

A main question was how the seven subscales would compare to personality scales that were purely empirically driven, their development constrained by their relation to other personality factors. The MAQ scales had similar cross-time stability (over a one-year span) as the HEXACO-H scale. With respect to internal consistency, the MAQ scales Fairness and Promise-Keeping fared the best across countries, better than the HEXACO-H and Big Six-H, while Truth-Telling and Compassion fared the worst, with lower than acceptable levels of alpha values. Fairness and Promise-Keeping were also the strongest MAQ scales across countries in informant-report data, in terms of internal consistency.

The MAQ scales were generally less unidimensional than the Big Six-H and HEXACO-H scales, with the exception of Beneficence and Non-Maleficence, which were comparable in their unidimensionality. For informant data, the Beneficence and Non-Maleficence scales also had the highest unidimensionality across countries; all other MAQ scales showed wide variation in their unidimensionality across countries (which was one contributor to the poor fit found in the CFAs and lack of measurement invariance found across countries).

The latent trait estimated using the Big Six-H and the latent trait estimated using the full MAQ inventory showed similar patterns regarding error of estimation across the latent trait— there was more measurement error for higher levels of the latent trait for both MAQ and Big Six-H, and this pattern also held for informant-rated MAQ. This could in part be due to the fact that the MAQ was predominantly designed to measure a set of minimum moral obligations that are agreed upon as morally significant across cultures, rather than supererogatory ones that reflect extreme or outstanding levels of virtue. Another explanation is the fact that the IPIP – the item pool from which the MAQ items were generated– lacks more neutral descriptors of ethical virtue. For example, a high-difficulty indicator of truth-telling might be telling the truth even when doing so would create stress or difficulties. The IPIP, in its emphasis on short, straight-forward items, lacks nuanced items like this, which may be helpful for creating indicators of higher levels of virtue.

A third way that this dissertation contributed to the assessment of ethical virtue was in its examination of the structure of ethical virtue across countries and different types of data. Across self- and informant-report data, models with more factors tended to have better fit in the training sets, but the downside of models with more factors, particularly factors that are highly correlated, is that estimation problems are more likely to occur in cross-cultural data, which is exactly what happened here. A model with two broad factors worked best cross-culturally in both self- and informant-report data (but did not have acceptable levels of fit cross-culturally; none of the models did). One of these factors was characterized by ethical virtues centered on avoiding violation of minatory ethical duties (Promise-Keeping, Truth-Telling, and Fairness), and another was focused on fulfilling

hortatory duties, which are more likely to be motivated primarily by social emotions such as sympathy (Compassion and Beneficence). The pattern of factor correlations roughly suggested that Non-Maleficence and Gratitude/Reciprocity were interstitial to these two domains, meaning that these ethical virtues may be a blend of the two factors. Both Non-Maleficence and Gratitude/Reciprocity have a clear impact on the well-being of individual others as the hortatory duties do, but they also have a strong rational basis informed by the more obligatory principles of fairness and fulfilling minimum expectations for behavior, so an argument could be made that these factors are interstitial from a rational as well as empirical basis.

In the informant data, the model that fit the best across countries was a two-factor model which dropped the items for Gratitude/Reciprocity entirely, and had two factors corresponding to what could be labeled justice (Promise-Keeping, Truth-Telling, and Fairness) and care (Beneficence, Non-Maleficence, and Compassion). This is consistent with what Gibbs (2013) suggests is at the core of the moral domain and moral development: moral sensibilities about the “right” and the “good.” Gibbs (2013) characterization of the “right” involves knowledge about moral responsibilities, fair-mindedness and respect for the autonomy of others, and his characterization of the “good” involves all forms of action intended to benefit or minimize the risk of harm to others. In the present data, it is difficult to determine whether a model distinguishing hortatory duties and minatory ones is most accurate, or one distinguishing care and justice, as results are different at the item-level than at the scale-level, and multiple MAQ scales had problems with multi-dimensionality.

The current results suggest factorial and scalar measurement invariance across countries cannot be assumed for either the self- or informant-report

version of the Moral Actions Questionnaire. This means that the content of the dimensions in each model may vary across countries, and any country-level mean differences observed in the dimensions should not be attributed to true mean differences in ethical virtues. Though the lack of measurement invariance implies that the current data cannot justify cross-cultural inferences, it is not necessarily a problem for the structural validity of the Moral Actions Questionnaire. It may be the case that the meaning of each of these ethical virtues truly is different cross-culturally, and an emic approach is most appropriate, where a different model of ethical virtue is derived given considerations specific to that country.

In general, the informant scales were much more highly correlated than the self-report scales, suggesting that people differentiate among these ethical virtues more for themselves than they do for others. Thus, although informant scales seem to be more advantageous for the purpose of forecasting, the self-report scales seem to be more differentiating. So both self- and informant-reports have their own utility for measuring the domain of ethical virtue.

The dimensions that were most independent across both self-report and informant data were Promise-Keeping and Compassion, which raises the question of whether there are underlying psychological processes arising in (relatively) independent variation in these two ethical virtues in particular. It seems plausible that social emotions such as empathy may be required for compassion, while self-regulatory abilities or a sense of duty may motivate the more minatory ethical virtues (centered on Promise-Keeping). However, another key distinction between these virtues lies in the consequence that ensues when one acts in accordance with or violates them: Violating minatory ethical virtues such as keeping one's promises is more likely to result in sanctions than failing to act compassionately in a given

instance. As the relevance of ethical virtues is largely in their impact on others, it is plausible that what separates these domains are differences in the response from one's environment when the virtue is acted out: punishment or criticism, in the case of violations of promise-keeping, and praise in the case of compassion.

Fourth, the results suggested some clear areas of needed improvement for the Moral Actions Questionnaire. The Truth-Telling items, though internally consistent in the U.S. samples, showed different patterns in their relationship to each other across countries (see Appendix A). This may be due to the fact that one of the items chosen for this scale used an idiom that may not work well in a cross-cultural context (“stretches the truth”) and others implied varied reasons for deception or telling the truth— items included “is not good at deceiving other people,” and “tells other people what they want to hear.” The Compassion scale and Gratitude/Reciprocity scale each formed two dimensions, so future versions of these scales should diversify the content so that these scales more evenly represent relevant behaviors to this domain. Alternatively, each scale could be split up into two separate scales, each with their own content. In addition, all scales should be longer in order to raise their alpha to standard levels of acceptability.

The IRT analysis suggested that more items that are difficult to endorse should be added, so that the latent trait is more precisely estimated for those higher in ethical virtue. And finally, future versions of this scale should not rely solely on IPIP items, as one limitation, particularly for the Truth-Telling scale, was the number of well-performing items available from the IPIP for inclusion in each scale. In general, the IPIP is currently heavily used for the development of personality scales, and one main lesson from this dissertation research is that heavy reliance on this item pool may lead to difficulties with respect to



generalizability to international samples with less familiarity with idioms specific to the American-English context, as well as difficulties with respect to IRT criteria.

Fifth, the current results had some implication for theories about how ethical virtues develop, which can be tested more stringently in the future using longitudinal data and improved scales. The validity indicator with the strongest and most consistent relation to the ethical virtue scales across countries was guilt proneness (but not shame proneness). Both empirical literature on moral development and philosophical theories regarding how ethical virtues develop converge on the role of guilt in moral behavior. Cohen et al. (2011) found that the tendency to experience guilt (defined both as a tendency to engage in reparative behavior after doing something wrong, and as a tendency to experience a negative emotional response to acting immorally in situations without a public embarrassment component) was positively related to Honesty/Humility from the HEXACO model, and negatively related to manipulative work behaviors as measured by the Self Reported Negotiation Strategies II Scale (SINS-II) (Cohen et al., 2011). In addition, both Aristotle and Plato suggest that shame is “a semi-virtue of the learner” (Burnyeat, 1980)– in order to be amenable to moral education, children must feel a sense of internal pain rather than pain due solely to external consequences. In the *Rhetoric*, Aristotle’s definition of shame is more consistent with the operationalization of guilt proneness rather than shame proneness on the GASP scale, as he defines it as “pain or disturbance in regard to bad things, whether present, past, or future, which seem likely to involve us in discredit...we feel shame at such bad things as we think are disgraceful to ourselves or to those we care for. These evils are, in the first place, due to moral badness.” Aristotle (350 BCE) then goes on to list a number of acts that one should feel

shame about, including acting cowardly, wronging people about money, and making profit in disgraceful ways. From these examples it is clear that Aristotle does not restrict his definition of shame to acts that are publicly observed. Instead, his conception of shame encompasses feeling internally pained due to any wrongful act.

A final note about compassion: according to many ethical theorists, compassion is the most fundamental of the ethical virtues, but it is also considered a nascent affective tendency by Mencius that needs to be cultivated in order to grow into the full virtue of benevolence (Flanagan, 2014). In its nascent form, compassion is felt in response to particular instances of suffering, but not when encountering all instances of suffering where a compassionate response would be appropriate. Mencius describes an example of how full benevolence may be cultivated from the nascent sprout of compassion in a king who is indifferent to the suffering of his people, but compassionate to the suffering of an ox about to be slaughtered. In passage IA7—from the translation by Eno (2016)—Mencius asks this king: “why would one accept that Your Majesty’s kindness could extend even to the birds and beasts, but its works could not extend to the people?” In this passage Mencius demonstrates how the felt sense of compassion in response to the ox may be cultivated so that the king feels compassion toward his people, and ultimately, feels a sense of compassion toward all suffering, directly observed or imagined. This example demonstrates that the felt sense of compassion may not be a full virtue in and of itself. This could be part of the reason the Compassion scale tended to show weaker relations to validity outcomes than the other scales; its possible that the Compassion scale was not capturing a full virtue, but rather a prerequisite to developing virtue. Future research that examines compassion

longitudinally in conjunction with other values and beliefs can lend more insight as to whether compassion plays a role in the development of other ethical virtues.

## CHAPTER V

### CONCLUSION

The primary purpose of this dissertation was to evaluate the strengths and weaknesses of the Moral Actions Questionnaire, relative to models of ethical virtue represented in already-existing personality inventories. An inventory that uses a rationale-based approach to more comprehensively represent the moral domain has the potential to remedy the fact that historically, moral content was partially excluded from personality inventories. Though the present findings highlight multiple areas of improvement for the Moral Actions Questionnaire, particularly in a cross-cultural context, they also suggest that a more comprehensive measure of ethical virtue does indeed predict important outcomes above and beyond what ethical virtue measures represented in already-existing personality inventories predict. In addition, the differentiation between conceptually distinct ethical virtues offered by the Moral Actions Questionnaire can inform investigations into what broad dimensions characterize moral functioning, and help guide theoretical perspectives on what leads to individual differences in these dimensions.

## APPENDIX A

### A CLOSER LOOK AT TRUTH-TELLING AND COMPASSION

Table A.25. Correlations between Truth-Telling items in self-report data

Country	52 & 85	52 & 100	52 & 104	85 & 100	85 & 104	100 & 104
Kenya	.05	-.05	-.20	.13	.04	.03
India	.04	-.18	-.02	.06	.07	.13
HKS	.19	-.07	-.32	-.03	-.01	.21
USA	.29	-.08	-.28	-.19	-.26	.25

Table A.26. Correlations between Compassion items in self-report data

Country	88 & 71	88 & 67	88 & 42	71 & 67	71 & 42	67 & 42
Kenya	.37	.02	-.05	-.04	-.11	.22
India	.34	-.03	.01	-.08	.05	.26
HKS	.27	-.09	-.16	-.08	.00	.10
USA	.39	-.16	-.26	-.26	-.25	.29

TRUTH-TELLING ITEMS: BFI100- tells other people what they want to hear; BFI104- tries to fool others; BFI52- is not good at deceiving other people; BFI85- seldom stretches the truth.

COMPASSION ITEMS: BFI42- looks down on any weakness; BFI67- likes the idea that only the strong should survive; BFI71- hates to see anyone suffer, even my worst enemy; BFI88- reacts compassionately to difficult people.

## APPENDIX B

### EXPLORATORY FACTOR ANALYSIS RESULTS

Table B.27. Results from item-level Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Student/Informant sample (N=842) (which informed structures tested in international samples)- two-factor solution

Subscale	Item	F1	F2	
BENEFCENCE	Inf37	.02	.67	goes out of his/her way to cheer up people who appear down.
BENEFCENCE	Inf44	-.12	.64	is sensitive to the needs of others.
BENEFCENCE	Inf21	.24	-.24	doesn't waste time with others' troubles.
BENEFCENCE	Inf28	.26	-.36	feels little concern for others.
COMPASSION	Inf25	.21	.72	reacts compassionately to difficult people.
COMPASSION	Inf41	.23	.78	hates to see anyone suffer, even his/her worst enemy.
COMPASSION	Inf16	.15	-.33	likes the idea that only the strong should survive.
COMPASSION	Inf32	.30	-.38	looks down on any weakness.
NONMAL	Inf26	-.14	.50	makes a point of trying not to hurt others in pursuit of his/her goals.
NONMAL	Inf31	.12	.66	prevents himself/herself from saying mean things.
NONMAL	Inf15	.13	-.50	is sometimes rude to others.
NONMAL	Inf42	.47	-.22	does things out of revenge.
FAIRNESS	Inf17	-.15	.58	has always been completely fair to others.
FAIRNESS	Inf40	-.40	.13	would never take things that aren't his/hers.
FAIRNESS	Inf24	.83	.11	cheats to get ahead.
FAIRNESS	Inf33	.69	-.13	takes advantage of others.
TRUTH	Inf10	-.04	.25	is not good at deceiving other people.
TRUTH	Inf29	-.14	.17	seldom stretches the truth.
TRUTH	Inf36	.54	-.12	tries to fool others.
TRUTH	Inf45	.57	.40	tells other people what they want to hear.
PROMISE	Inf22	-.44	.26	honors all the vows he/she has made.
PROMISE	Inf38	-.53	.13	follows through on his/her commitments.
PROMISE	Inf27	.74	.01	breaks his/her promises.
PROMISE	Inf43	.62	.07	shirks his/her duties.
GRAT/RECIP	Inf23	-.32	.41	expresses thanks to those who care about him/her.
GRAT/RECIP	Inf34	-.23	.45	is never too busy to help a friend.
GRAT/RECIP	Inf18	.54	-.23	neglects to thank others for their help.
GRAT/RECIP	Inf39	.53	-.18	doesn't see the need to acknowledge those who are good to him/her.

Table B.28. Results from item-level Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Student/Informant sample (N=842) (which informed structures tested in international samples)- three-factor solution

Subscale	Item	F1	F2	F3
BENEFCENCE	Inf37	.64	-.06	.01
BENEFCENCE	Inf44	.59	-.15	-.08
BENEFCENCE	Inf21	.01	-.03	.55
BENEFCENCE	Inf28	-.15	.05	.48
COMPASSION	Inf25	.60	.16	-.10
COMPASSION	Inf41	.65	.16	-.08
COMPASSION	Inf16	.00	-.17	.69
COMPASSION	Inf32	-.11	.02	.62
NONMAL	Inf26	.43	-.12	-.14
NONMAL	Inf31	.50	.13	-.21
NONMAL	Inf15	-.29	-.02	.42
NONMAL	Inf42	-.08	.27	.41
FAIRNESS	Inf17	.56	-.20	-.03
FAIRNESS	Inf40	.15	-.35	-.08
FAIRNESS	Inf24	.18	.56	.41
FAIRNESS	Inf33	-.02	.45	.43
TRUTH	Inf10	.19	-.02	-.10
TRUTH	Inf29	.21	-.18	.05
TRUTH	Inf36	-.01	.33	.39
TRUTH	Inf45	.32	.46	.06
PROMISE	Inf22	.42	-.57	.22
PROMISE	Inf38	.33	-.67	.28
PROMISE	Inf27	-.09	.72	.00
PROMISE	Inf43	.07	.47	.20
GRAT/RECIP	Inf23	.45	-.33	-.01
GRAT/RECIP	Inf34	.44	-.24	-.06
GRAT/RECIP	Inf18	-.16	.38	.31
GRAT/RECIP	Inf39	-.09	.34	.36

Table B.29. Results from Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Community sample (n=538)- two-factor solution

		F1	F2	
BENEFICENCE	T91	-.36	.14	goes out of his/her way to cheer up people who appear down.
BENEFICENCE	T112	-.58	.00	is sensitive to the needs of others.
BENEFICENCE	T26	.59	.13	doesn't waste time with others' troubles.
BENEFICENCE	T76	.75	.30	feels little concern for others.
COMPASSION	T103	-.47	-.09	reacts compassionately to difficult people.
COMPASSION	T79	-.35	-.04	hates to see anyone suffer, even his/her worst enemy.
COMPASSION	T74	.72	.37	likes the idea that only the strong should survive.
COMPASSION	T45	.49	.09	looks down on any weakness.
NONMAL	T116	-.31	.18	makes a point of trying not to hurt others in pursuit of his/her goals.
NONMAL	T123	-.23	.21	prevents himself/herself from saying mean things.
NONMAL	T108	.28	-.14	is sometimes rude to others.
NONMAL	T62	.30	-.16	does things out of revenge.
FAIRNESS	T95	.18	.61	has always been completely fair to others.
FAIRNESS	T40	-.11	.33	would never take things that aren't his/hers.
FAIRNESS	T102	.13	-.38	cheats to get ahead.
FAIRNESS	T83	.32	-.28	takes advantage of others.
TRUTH	T55	-.12	.16	is not good at deceiving other people.
TRUTH	T100	-.03	.32	seldom stretches the truth.
TRUTH	T120	.09	-.14	tries to fool others.
TRUTH	T124	.17	-.34	tells other people what they want to hear.
PROMISE	T119	.13	.65	honors all the vows he/she has made.
PROMISE	T86	.33	.72	follows through on his/her commitments.
PROMISE	T97	-.25	-.84	breaks his/her promises.
PROMISE	T104	-.02	-.51	shirks his/her duties.
GRAT/RECIP	T96	-.28	.26	expresses thanks to those who care about him/her.
GRAT/RECIP	T106	-.14	.33	is never too busy to help a friend.
GRAT/RECIP	T78	.49	-.11	neglects to thank others for their help.
GRAT/RECIP	T92	.22	-.34	doesn't see the need to acknowledge those who are good to him/her.



Table B.30. Results from Exploratory Factor Analysis with promax rotation using maximum likelihood estimation in Community sample (n=538)- three-factor solution

Subscale	Item	F1	F2	F3
BENEFACTENCE	T91	.63	.23	.29
BENEFACTENCE	T112	.63	.03	.03
BENEFACTENCE	T26	-.50	.11	.09
BENEFACTENCE	T76	-.58	.26	.13
COMPASSION	T103	.53	-.03	.11
COMPASSION	T79	.41	.01	.09
COMPASSION	T74	-.39	.43	.36
COMPASSION	T45	-.32	.13	.22
NONMAL	T116	.40	.18	.03
NONMAL	T123	.27	.17	-.07
NONMAL	T108	-.12	-.03	.33
NONMAL	T62	-.24	-.09	.20
FAIRNESS	T95	.01	.51	-.02
FAIRNESS	T40	.11	.23	-.19
FAIRNESS	T102	.04	-.19	.47
FAIRNESS	T83	-.03	-.06	.63
TRUTH	T55	.00	.06	-.28
TRUTH	T100	-.07	.18	-.31
TRUTH	T120	.18	.03	.49
TRUTH	T124	.05	-.14	.53
PROMISE	T119	.06	.55	-.04
PROMISE	T86	-.11	.59	-.03
PROMISE	T97	.12	-.63	.23
PROMISE	T104	.10	-.31	.40
GRAT/RECIP	T96	.31	.21	-.10
GRAT/RECIP	T106	.30	.31	.06
GRAT/RECIP	T78	-.36	-.03	.25
GRAT/RECIP	T92	-.23	-.26	.16

## APPENDIX C

### RELIABILITY OF RATERS FOR SIX MAQ DOMAINS

Table C.31. Standardized alpha values for 3 raters' ratings of 2458 IPIP items' relevance to each of the original six MAQ domains

Moral Domain	Standardized Alpha
Promise-keeping	.664
Truth-telling	.699
Reciprocity	.649
Non-maleficence	.714
Beneficence	.733
Fairness	.634

## REFERENCES CITED

- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), i.
- Aristotle. (350 BCE). *The Rhetoric*. Retrieved from <http://classics.mit.edu/Aristotle/rhetoric.2.ii.html> (Translated by W. Rhys Roberts)
- Ashton, M., & Lee, K. (2004). *HEXACO Personality Inventory–Revised [Database Record]*. (Retrieved from PsycTESTS) doi: <http://dx.doi.org/10.1037/t03138-000>
- Ashton, M., Lee, K., & De Vries, R. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, 18(2), 139–152.
- Bok, S. (1996). *Common values*. University of Missouri Press: Columbia, MO.
- Bollen, K. A. (2009). *Structural equations with latent variables*. Wiley.
- Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press.
- Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature*, 425(6955), 297–299.
- Burnyeat, M. (1980). Aristotle on learning to be good. In A. O. Rorty (Ed.), *Essays on Aristotle's ethics* (p. 69-92). Oakland, CA: University of California Press.
- Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology*, 101(4), 831.
- Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, 100(5), 947.
- Costello, C., Srivastava, S., & Saucier, G. (2017). Stability and change in the Big Five and Big Six: New tests of the maturity and cumulative continuity principles. *PsyArXiv Preprints*. Retrieved from <https://psyarxiv.com/b36tv>

- De Raad, B., & Hoskens, M. (1990). Personality-descriptive nouns. *European Journal of Personality*, *4*, 131–146.
- DeSalvo, K. B., Fan, V. S., McDonell, M. B., & Fihn, S. D. (2005). Predicting mortality and healthcare utilization with a single question. *Health Services Research*, *40*(4), 1234–1246.
- Doris, J. (2002). *Lack of character: Personality and moral behavior*. Cambridge, UK: Cambridge University Press.
- Elster, J. (2007). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press.
- Eno, R. (2016). *Mencius: an online teaching translation*. Robert Eno. Retrieved from <http://www.indiana.edu/~p374/Mengzi.pdf>
- Findlay, J. (1961). *Values and intentions: A study in value-theory and philosophy of mind*. London: George Allen & Unwin.
- Flanagan, O. (2014). *Moral sprouts and natural teleologies: 21st century moral psychology meets classical Chinese philosophy*. Milwaukee, WI: Marquette University Press.
- Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The prospects for a personality-based perspective on morality. *Social and Personality Psychology Compass*, *8*(4), 178–191.
- Gibbs, J. C. (2013). *Moral development and reality: Beyond the theories of Kohlberg, Hoffman, and Haidt*. Oxford University Press.
- Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychological Bulletin*, *136*(3), 351.
- Goldberg, L. R. (1990). An alternative “description of personality”: the Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229.
- 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*(2), 366–385.
- Hill, P. L., & Roberts, B. W. (2010). Propositions for the study of moral personality development. *Current Directions in Psychological Science*, *19*(6), 380–383.

- Iurino, K. (2012). *Predicting moral behavior from beliefs* (Unpublished master's thesis).
- John, O., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, 2, 102–138.
- Kant, I. (2001). Morality and rationality. In G. Sher (Ed.), *Moral philosophy: Selected readings, second edition*. Belmont, CA: Wadsworth/Thompson Learning.
- King, L. A., Heintzelman, S. J., & Ward, S. J. (2016). Beyond the search for meaning: A contemporary science of the experience of meaning in life. *Current Directions in Psychological Science*, 25(4), 211–216.
- Lee, K., & Ashton, M. C. (2013). *The H factor of personality: Why some people are manipulative, self-entitled, materialistic, and exploitive— and why it matters for everyone*. Waterloo, Canada: Wilfrid Laurier University Press.
- Marks, N. F., & Song, J. (2009). Compassionate motivation and compassionate acts across the adult life course: Evidence from US national studies. *The Science of Compassionate Love: Theory, Research, and Applications*, 121–158.
- Nagel, T. (1991). *Mortal questions*. Cambridge: Cambridge University Press.
- Nussbaum, M. (1996). Compassion: The basic social emotion. *Social Philosophy and Policy*, 13(1), 27–58.
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164–172.
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Revelle, W. (2017). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from <https://CRAN.R-project.org/package=psych> (R package version 1.7.8)
- Roberts, B. W., Bogg, T., Walton, K. E., Chernyshenko, O. S., & Stark, S. E. (2004). A lexical investigation of the lower-order structure of Conscientiousness. *Journal of Research in Personality*, 38(2), 164–178.

- Robitzsch, A., Kiefer, T., & Wu, M. (2018). TAM: Test analysis modules [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=TAM> (R package version 2.9-35)
- Ross, D. (2003). *Right and the good*. Oxford, UK: Oxford University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Saucier, G. (2000). Isms and the structure of social attitudes. *Journal of Personality and Social Psychology*, *78*(2), 366.
- Saucier, G. (2003). Factor structure of English-language personality type-nouns. *Journal of Personality and Social Psychology*, *85*(4), 695.
- Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. *Journal of Personality*, *77*(5), 1577–1614.
- Saucier, G. (2017). Personality, character, and cultural differences: Distinguishing enduring-order versus evolving-order cultures. In A. Church (Ed.), *Personality across cultures*. Santa Barbara, CA: Praeger.
- Saucier, G., Kenner, J., Iurino, K., Bou Malham, P., Chen, Z., Thalmayer, A. G., ... others (2015). Cross-cultural differences in a global survey of world views. *Journal of Cross-Cultural Psychology*, *46*(1), 53–70.
- Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, *76*(4), 613.
- Saucier, G., Thalmayer, A. G., & Bel-Bahar, T. S. (2014). Human attribute concepts: Relative ubiquity across twelve mutually isolated languages. *Journal of Personality and Social Psychology*, *107*(1), 199–216.
- Schlenker, B. R., Weigold, M. F., & Schlenker, K. A. (2008). What makes a hero? The impact of integrity on admiration and interpersonal judgment. *Journal of Personality*, *76*(2), 323–355.
- Schnell, T. (2009). The sources of meaning and meaning in life questionnaire: Relations to demographics and well-being. *The Journal of Positive Psychology*, *4*(6), 483–499.
- Schopenhauer, A. (2014). *On the basis of morality*. Retrieved from <https://www.gutenberg.org/files/44929/44929-h/44929-h.htm> (original German publication 1840; translated EBook released February 16, 2014)

- Schwartz, S. H. (2009). Draft Users Manual: Proper Use of the Schwarz Value Survey, Version 14 January 2009, compiled by Romie F. Littrell. Auckland, New Zealand: Centre for Cross-cultural comparisons. Retrieved from <http://www.crossculturalcentre.homestead.com>
- Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures: Taking a similarities perspective. *Journal of Cross-cultural Psychology, 32*(3), 268–290.
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment, 23*(4), 995.
- Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*(2), 281–300.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*(5), 1202–1216.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one, 11*(3), e0152719.