



# Intrinsic and extrinsic religiousness: genetic and environmental influences and personality correlates

Thomas J Bouchard Jr, Matt McGue, David Lykken and Auke Tellegen

University of Minnesota, Minneapolis, USA

This report presents findings for the Intrinsic (IR) and Extrinsic (ER) religiousness scales from the Minnesota Study of Twins Reared Apart. The scales were shown to be internally consistent, sufficiently distinct from the scales of the California Psychological Inventory and the Multi-dimensional Personality Questionnaire and unrelated to a number of measures of response style to justify treating them as distinct traits. The I scales also showed considerable evidence of construct validity in its correlations with religious fundamentalism and authoritarianism as assessed by the MMPI and Altemeyer's Right-Wing Authoritarianism scale. Data on IR and ER from 35 pairs of monozygotic twins reared apart (MZA) and 37 pairs of dizygotic twins reared apart (DZA) were fitted to a biometric model and demonstrated significant heritability (0.43 and 0.39), with a model containing genetic plus environmental factors fitting significantly better than a model containing only an environmental component. Twin similarity could not be explained by placement on a self-reported measure of family Moral Religious Emphasis as measured by the Family Environment Scale.

Keywords: twins, religiousness, personality

## Introduction

According to EO Wilson, 'The predisposition to religious belief is the most complex and powerful force in the human mind and in all possibility an inerradical part of human nature.'<sup>1</sup> This evolutionary theme has now been explored by a number of scholars,<sup>2,3</sup> although it appears to have been largely ignored by evolutionary psychologists.<sup>4,5</sup> Wilson's claim is, of course, a reiteration of similar assertions made over the millennia by thoughtful people. Nevertheless, in spite of the recognition of its pervasiveness, complexity and importance, religiousness as a psychological trait has been neglected by modern psychologists relative to most other psychological constructs. It has been relegated, perhaps not to the fringes, but to the sidelines. This is surprising given the importance attached to religiousness by both William James,<sup>6</sup> a founding father of modern psychology and the author of one of the discipline's most famous textbooks,<sup>7</sup> and Gordon Allport,<sup>8</sup> the father of modern personality theory. The most recent personality textbooks to cross our desks<sup>9,10</sup> do not contain the terms religion or religiousness in their indexes. Thomas and Carver<sup>11</sup> report similar findings based on their examination of 60 textbooks devoted to child and adolescent devel-

opment. The failure to cite religiousness as an important variable in major textbooks is also surprising because religiousness is a powerful inverse predictor of most of the debilitating conditions psychologists purport to be interested in understanding and alleviating – alcoholism, drug abuse, mental health, physical health, and delinquency.<sup>12–14</sup> These empirical facts alone should make religiousness a prime candidate for extensive scientific analysis.

Our interest in religiousness as a psychological trait, however, flows from Wilson's claim<sup>1</sup> that light may be cast on the nature of this trait if we pay attention to the sociobiology of religion: 'By traditional methods of reduction and analysis science can explain religion but cannot diminish the importance of its substance' (p 172). Until quite recently most psychologists did not even question the assumption that religious attitudes and beliefs were largely shaped by parental influences. Sociologists research the 'inheritance' of religiousness without even mentioning the possibility that genes may be involved.<sup>15</sup> Even behavioral geneticists who have tested the assumption of common environmental influence have been misled. Plomin,<sup>16</sup> for example, asserted: 'Religiosity, for example, shows no genetic influence and is thus thought to be due to environmental influence' (p 97). This claim was made on the basis of data gathered on young people who had not yet 'left the nest'. It is now known that the effects of family environment diminish after children leave their homes.<sup>17,18</sup> Wilson has recently made very clear what he means by traditional methods of reduction

Correspondence: Thomas J Bouchard Jr, Department of Psychology, Elliott Hall, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455, USA. Fax: 612 6262079; E-mail: [bouch001@c.umn.edu](mailto:bouch001@c.umn.edu)

Received 29 April 1999; accepted 4 May 1999

and analysis. Specifically they include replication, mensuration, parsimony, heuristics and consilience.<sup>19</sup> We believe a few comments on the first four as they apply to religiosity would be worthwhile. The last, consilience, is unquestionably relevant but beyond the scope of this paper.

### Replication

It is now reasonably well established that variation in religiosity is, contrary to widespread belief, moderately influenced by genetic factors.<sup>20</sup> This finding throws all studies of the influence of environmental variables, based on studies of biological families, into serious doubt, as such designs assume rather than test for environmental influences. The finding of genetic influence on any trait raises an entire array of new questions, for example, what is, a) the degree of additive and/or nonadditive genetic influence, b) the nature of the environmental influences (shared or unshared), c) the role of genes and environment on developmental change, d) the influence of assortative mating, etc. We already have some inklings regarding these influences, but it should be clear that such findings must also be subjected to constructive replication. Constructive replication involves the use of alternate designs and tools of measurement to demonstrate that a finding is independent of the unique features of a particular design and measurement tool.<sup>21</sup> This study addresses both of these issues as we utilize twins reared apart, a rare design, and employ the Intrinsic (IR) and Extrinsic (ER) measures of religiosity, scales not previously used in behavior genetic studies. Additional designs, not involving twins, especially studies of unrelated individuals reared together as children and followed up as adults, would be very desirable. That genetic factors influence religiosity is a sufficiently counter-intuitive finding that it will be accepted as an empirical fact only after it has been constructively replicated numerous times.

### Mensuration

Constructive replication requires the existence of adequate measuring instruments. The problem of adequate measurement plagues the study of religiosity. More and better measuring instruments are indispensable for advance to occur. When measures of religion/religiosity are included in studies, they often tend to be simple nonfocal variables such as denominational membership, frequency of attendance at religious services, or a rating of the importance of religion to the individual. More specific scales often focus only on Christian orthodoxy rather than attempt to capture a common

component of the diverse belief systems in most of the populations actually being studied.<sup>22</sup> Better and more diverse instrumentation is sorely needed.

The study reported in this paper examines genetic and environmental influences on Intrinsic (IR) and Extrinsic (ER) religiosity. The distinction between these two measures is a fundamental one in the domain of the psychology of religion.<sup>13,23</sup> The distinction was best summed up by Allport and Ross<sup>24</sup> who stated that 'the extrinsically motivated person uses his religion, whereas the intrinsically motivated person lives his religion' (p 434). The IR and ER scales distinguish between instrumental values (Extrinsic Religiosity) as a mode of conduct and terminal values (Intrinsic Religiosity) as an end state of existence. This distinction also underlies Rokeach's widely used instrument, the Value Survey,<sup>25</sup> in which instrumental and terminal values constitute separate lists that are ranked independently of each other. When we decided to include a measure of religiosity in the Minnesota Study of Twins Reared Apart (MISTRA), a number of colleagues suggested to us that many measures of religiosity largely reflect response sets – socially desirable responses and acquiescence – and/or perhaps a manipulative/social use of religion, not 'true' or 'intrinsic' religiosity. A search of the literature led us to the Age Universal Religious Orientation Scale (AUROS).<sup>26</sup> This instrument is a modified version of the Allport and Ross I–E scales.<sup>24</sup> Gorsuch and Venable revised the I–E scale to allow its use with children and young adolescents by rewriting items to match a fifth grade reading level. This latter feature was important to us because of the extremely wide range of educational background and reading ability among the participants in MISTRA. For example, numerous participants have found the Study of Values<sup>27</sup> very difficult and confusing because of the complexity level of the questions and the college level content.

There is also good evidence that social desirability is not a problem with various measures of IR and ER.<sup>13,23,28</sup> Our twins had already completed the California Psychological Inventory (CPI) and the Multidimensional Personality Questionnaire (MPQ). Both instruments contain a number of validity scales. This allowed us to evaluate directly the possible role of social desirability and acquiescence in this sample, as well as examine the IR and ER scale correlations with well established personality inventories. On the CPI the Good Impression (Gi) scale measures the tendency to paint an excessively positive picture of one's self. The Dicken Social Desirability scale (DSD) is also a measure of social desirability and the Dicken Acquiescence scale (DAC) is a measure of yea saying. Strong positive correlations between these scales and the IR and ER

scales would throw doubt on the usefulness of the AUROS as a measure of religiousness. The MPQ contains a number of validity scales one of which, Unlikely Virtues, also assesses the social desirability response set.

### Parsimony

One reason psychologists have not studied religiousness in more detail is because many of them believe that it is already subsumed by major personality traits. The evidence simply does not support this conclusion. The correlations between measures of religiousness and personality are generally small<sup>20</sup> and, when they exist, different personality measures correlate differentially with different measures of religiousness. Altemeyer,<sup>29</sup> for example, has shown that IR and ER correlate differentially with the Right Wing Authoritarianism scale (RWA). IR correlates 0.36 and 0.41 with RWA in student ( $n = 406$ ) and parent ( $n = 549$ ) samples. The correlations with ER were  $-0.10$  and  $-0.09$ . He presents intercorrelations between a wide range of religiousness measures (as well as some scales related to religious attitudes and beliefs created for his studies) and RWA. All of the validity coefficients for IR are in the 0.55 to 0.65 range, whereas the correlations for ER are negative and in the  $-0.15$  to  $-0.30$ , range confirming the conclusion that IR does indeed have construct validity and that ER measures a different kind of religiousness at best. The correlations found in the parent and student samples are essentially the same, thus providing a replication across generations. There is some suggestion that correlations between religiousness measures and personality differ somewhat from one religious group to another.<sup>30</sup> Such findings, however, may simply reflect the use of small samples and require replication with much larger samples.

There is no support for the argument that religiousness should be excluded from the armamentarium of individual difference measures on the grounds of parsimony. On the contrary, the lack of citation of religiousness in most textbooks shows that the question of parsimony has hardly begun to be addressed. The location of various religiousness measures in the domain of psychological variables necessary to cover comprehensively the entire array of meaningful human individual differences has yet to be determined and deserves a great deal more attention.

### Heuristics

The issue of heuristics has already been touched upon. Religiousness is an important construct because it is significantly and substantively related

to many real-life conditions of intense interest to social scientists. Those who see religiousness as a positive trait emphasize its predictive significance for mental and physical health behavior. On the other hand, those who see it as a malevolent force relate it to such variables as Right Wing Authoritarianism, Dogmatism and other 'negative attitudes' and largely ignore its positive aspects.<sup>29,31</sup> Religiousness appears to be a psychologically complex trait that lies at the center of a nexus of correlations involving, among other things, important life style choices influencing physical health, mental states influencing mental health and important attitudinal variables (some positive, some negative). A thorough understanding of the various facets of religiousness and their causal antecedents would move us a long way toward understanding the genesis of many real life problems.

This study also allows a test of an additional hypothesis about IR and ER. In a previous study of genetic influence on measures of religiousness (not the IR and ER scales used in this study), using MZA and DZA twins, we found that some of the DZA twin correlations were near zero suggesting the trait might involve a configural, nonadditive combination of genetic effects.<sup>32</sup> We predict similar findings here.

## Method

### Subjects

The reared-apart twins were participants in MISTRA between the years of 1979 and 1998. Details of their recruitment and testing are found elsewhere.<sup>33,34</sup> To ensure comparable samples of adult participants across publications, no participants under 18 years of age are included in the sample used in this paper. Table 1 presents the descriptive characteristics of the sample. Note that the AUROS and RWA scales were mailed out to many of the twins (see below), so the figures given in Table 1 described the sample at the time of intake assessment only.

Females are over-represented as they are in most twin studies.<sup>35</sup> The twins are primarily middle-aged adults, and have experienced varying degrees of separation and contact. Zygosity for all pairs was determined by the Minneapolis Memorial Blood Bank (MMBB) which has changed the genetic systems assessed for determination of twin type in recent years. Until 1994, MMBB never used fewer than nine protein markers (blood groups, serum proteins and enzymes). MMBB currently uses six blood group markers and three DNA polymorphisms. The probability that a DZA pair would be concordant on all markers and thus misclassified as an MZA pair is less than 0.001. One twin pair was diagnosed on the basis of observation by study

Table 1 Means and standard deviations for age at assessment, measures of Separation and Contact and percentage of females for MZA and DZA twin samples

	Age (years)	% Females	Time Together Prior to Separation (days)	Total Contact Time (weeks)	Time Apart (months)
MZA (n = 35)					
Mean	41.9	65.7	151.1	68.4	415.7
SD	12.2		244.2	116.7	202.4
DZA (n = 37)					
Mean	47.2	56.8	287.1	46.2	528.3
SD	14.2		390.9	55.1	171.4

investigators because their blood sample was unusable upon reaching the MMBB. Whenever possible, spouses and partners and in some cases children and friends of twins participate in the study thus providing information on assortative mating and a large sample for age and sex correction of the data (see below) and examination of the correlations between instruments based on a sample of adequate size.

### Instruments

Among the many instruments completed by the participants in MISTRA the following were used in this paper. The 480-item version of the California Psychological Inventory (CPI),<sup>36</sup> the Minnesota Multiphasic Personality Inventory (MMPI),<sup>37</sup> the Multi-dimensional Personality Questionnaire (A Tellegen, 1982, unpublished manuscript), the Family Environment Scale,<sup>38</sup> a modified version of the AUROS and the Right-Wing Authoritarian scale (RWA).<sup>29</sup> Because of factors such as age, infirmities, time available for assessment, not all twins were able to complete all components of the assessment. The AUROS and the RWA were incorporated into the assessment battery late in the study and at different times. At the time of incorporation they were mailed to the previous twin participants, who we were confident would complete them in a reliable fashion. Both scales have been administered to subsequent twins as part of the regular assessment. We modified the AUROS somewhat to accommodate religions other than those from a Judeo-Christian heritage. The standard response format is a five point Likert scale running from strongly disagree to strongly agree. We modified this format to be consistent with other instruments in our battery (-4 = Very strongly disagree, -3 = Strongly disagree, -2 = Moderately disagree, -1 = Slightly disagree; 0 = Exactly and precisely neutral etc to +4 = Very strongly agree). Maintaining a consistent response format over a number of different instruments reduces confusion among the many participants in our study who have never before completed psychological inventories.

Consistent with previous procedures we provided the participants with the option of responding 'not

applicable' (n/a). Respondents used the n/a response rather liberally and it appeared that we would lose many cases due to there being too few items on the already brief scales. Upon questioning participants in MISTRA, we found that they almost always used n/a to mean not sure one way or the other. This is the meaning of the middle or neutral response according to Gorsuch and Venable.<sup>26</sup> Non-responses are a common problem when using the I-E scales with samples containing a sizeable proportion of non-religious participants. The standard procedure is to count non responses as neutral responses.<sup>23,29</sup> We followed that procedure here, scoring n/a responses as 0. Intrinsic (IR) or Extrinsic (ER) scores were derived following Gorsuch and Venable,<sup>26</sup> taking into account the item keying correction reported by Gorsuch.<sup>13</sup> The revised items used in this study are shown in Table 2. Items 1, 5, 6, 8, 10, 11, 15, 18 and 20 constitute the IR scale, and items 2, 3, 4, 7, 9, 12, 13, 14, 16, 17, and 19 constitute the ER scale.

### Analytic procedures

#### Age and sex

Scores on each of the scales were corrected for age and sex using procedures described by McGue and Bouchard<sup>39</sup> based on all the scores available (n = 253).

#### Analysis of the influence of separation and contact

Reared-apart twin studies are not truly randomized experiments because the twins vary in their degree of separation and their degree of contact. As indicated in Table 1, these twins were separated at various ages and had varying degrees of contact prior to participation in the MISTRA assessment. The similarity of twins reared apart on various psychological traits has sometimes thought to be due to contact between the twins.<sup>40,41</sup> In order to test this hypothesis we gathered relevant data. The separation/contact measures in Table 1 were based on detailed interviews with both twins and a review of

Table 2 Varimax factor loadings of the Revised AUROS items for a two factor solution (n = 253). Loadings whose absolute value was less than 0.30 are omitted

Item number, text of item and scale on which the item is scored (IR or ER)	IR	ER
10. My religion is important because it answers many questions about the meaning of life. (IR)	0.77	
1. I enjoy reading about my religion. (IR)	0.75	
5. It is important to me to spend time in private thought and prayer. (IR)	0.74	
20. I would prefer to go to my place of worship <sup>a</sup> . (IR)	0.74	
6. I have often had a strong sense of God's presence. (IR)	0.73	
8. I try hard to live all my life according to my religious beliefs. (IR)	0.72	
15. My whole approach to life is based on my religion. (IR)	0.71	
18. Prayers I say when I'm alone are as important to me as those I say in my place of worship. (IR)	0.68	
3. It doesn't much matter what I believe as long as I am good. (ER)	-0.63	
11. I would rather join a religious study group (a group that studies the Bible, Koran, Torah or other religious text) than a social group at my place of worship (Church, Synagogue, Temple). (IR)	0.55	
9. What religion offers me most is comfort in times of trouble and sorrow. (ER)	0.53	0.36
12. Prayer is for peace and happiness. (ER)	0.44	
19. Although I believe in my religion, many other things are more important in life. (ER)	-0.39	
14. I go to my (Church, Synagogue, Temple) to spend time with my friends. (ER)		0.80
16. I go to my (Church, Synagogue, Temple) mainly because I enjoy seeing people that I know there. (ER)		0.79
2. I go to my place of worship (Church, Synagogue, Temple) because it helps me to make friends. (ER)		0.66
7. I pray mainly to gain relief and protection. (ER)		0.56
17. I pray mainly because I have been taught to pray. (ER)		0.50
4. Sometimes I have to ignore my religious beliefs because of what people might think of me. (ER)		0.42
13. Although I am religious, I don't let it affect my daily life. (ER)		
Percent Variance Accounted for	30.0	14.1

<sup>a</sup>Response format for item 20. a) a few times a year or less, b) once every month or two, c) two or three times a month, d) about once a week, e) more often than once a week, f) about once a day.

available records. The influence of these variables was evaluated by correlating each measure with twin pair absolute differences on the IR and ER scales. If twin contact accounts, at least in part, for twin similarity, then twin pair differences should be negatively correlated with indices of time spent together and positively correlated with time spent apart. Because the AUROS and the RWA were mailed to many of the twins after their participation in the assessment (long after in some cases and shortly thereafter in others) the measures of Total Contact and Time Apart are only rough approximations.

#### Analysis of the influence of self-reported family rearing environment

The Moral Religious Emphasis (MRE) scale of the FES is the theoretically most relevant possible predictor of I and E in this data set. If it is a trait-relevant causal variable and if the twins have undergone placement on this variable, then its influence may be important. The intraclass correlation for the age and sex-corrected scores on the MRE scale for the MZA twins (n = 68 pairs) is 0.32, suggesting a moderate amount of placement. The correlation between MRE and I for MISTRA participants reared by unrelated individuals is, however, only 0.10 (n = 127) suggesting at best a very weak causal influence. The correlation for participants reared by biological families is 0.53 (n = 42). The

biological correlation confounds genetic and environmental influences, whereas the adopted correlation is uncontaminated by genetic influence. To estimate the influence of MRE on the MZA intraclass correlation for IR, we need only multiply the magnitude of placement by the square of the relevant environmental correlations (0.32 × (0.10 × 0.10)). Under the hypothesis of no genetic influence the predicted MZA correlation for I would be 0.003. The correlation between MRE and E in both the adopted and biological groups is slightly negative and near zero. Placement, at least on MRE, does not appear to explain much twin similarity on IR and ER.

#### Model fitting

The quantitative genetic model assumes that observed phenotypic variance ( $V_p$ ) is a linear additive function of genetic ( $V_g$ ) and environmental ( $V_e$ ) variances. Symbolically,

$$V_p = V_g + V_e$$

As our twins have been reared apart, the environmental variance in the model represents residual variance not explained by genetic influence, ie nonshared environmental variance confounded with measurement error.

From quantitative genetic theory and under the assumption that all genetic effects are additive ( $V_a$  = variance due to additive gene effects), we can derive the expected covariance between any two

relatives as a function of the variance components given above. The expected covariances between the MZA and DZA twin pairs will be

$$\begin{aligned} \text{COV}_{(MZA)} &= V_a \\ \text{COV}_{(DZA)} &= 0.5V_a \end{aligned}$$

The general assumptions in the model are: (a) there is no genotype–environment correlation or interaction, (b) all genetic effects are additive effects, (c) mating is random with regard to the traits under study, and (d) no selective placement on trait relevant factors has occurred. We test the assumption of random mating below. Extensive discussions of the other assumptions in behavioral genetic designs are available elsewhere.<sup>42</sup>

Expected variances and covariances based on the model for the MZA and DZA twins were applied to the observed variances and covariances using a maximum likelihood estimation procedure in Mx.<sup>43</sup> One of the advantages of the model-fitting technique is that it provides the opportunity to test competing theoretical models. We fitted a full model and a reduced model to the twin covariances.

The full model included the additive genetic variance component and the unshared environment/measurement error variance component; the reduced model entailed dropping the additive genetic variance component. To measure overall fit of the model, we used the  $\chi^2$  test statistic with the criterion for rejection of models at the 0.05 probability level. In evaluating the relative fits of the various models, particularly those that could not be rejected on the basis of the  $\chi^2$  test statistic alone, the Akaike Information Criterion [AIC =  $\chi^2 - 2(df)$ ] was examined. AIC quantifies the information content of a model in terms of the joint criteria of fit and parsimony. In general, small  $\chi^2$  values from models with few free parameters lead to small AICs, representing maximum parsimony, whereas large  $\chi^2$  values from models with many parameters yield large AICs, representing lack of parsimony.<sup>44</sup>

## Results

The alpha reliabilities, based on 253 cases, were 0.88 for IR and 0.66 for ER. The IR items clearly form a coherent and internally consistent scale. The ER scale is much less internally consistent.

The correlation between IR and ER for all cases ( $n = 253$ ) was 0.04, confirming that in a religiously heterogeneous sample the measures are uncorrelated. A factor analysis of the items yielded three eigenvalues greater than 1.00 and a readily interpretable two factor solution, although there was a small number of items that loaded on both factors. The results of the two factor solution are given in

Table 2. The factor scores correlated 0.97 (IR) and 0.93 (ER) with the scales. We chose to analyze the standard scales rather than factor scores in order to allow a more direct comparison of our findings with those of other studies that have used the IR and ER scales. The means and standard deviations for the IR and ER scales, by zygosity, are shown in Table 3. The MZA and DZA twins do not differ significantly with respect to means or variances.

The correlations between absolute differences between members of each twin pair on IR and ER and separation and contact are given in Table 4 for both types of twin. None of the correlations is significant for the MZA twins for either measure. All three measures, however, correlated significantly with ER in the DZA sample and the correlations with IR, while not statistically significant, are similar. The correlations are all in the predicted direction. Time raised apart correlates positively with absolute difference and amount of contact correlates negatively. Total Contact Time (TCT) incorporates Time Together Prior to Separation (TTPS). Contact after Separation (TCT–TTPS) has no correlation with either IR (0.07) or ER (–0.04); consequently all the correlation is accounted for by the measure of early contact. Time Apart is correlated –0.10 with TTPS; consequently it contributes independently to the prediction of the difference score. The adjusted multiple  $r$  for predicting ER is 0.42. The intraclass correlations and 95% confidence intervals for the IR and ER scales for the MZA and DZA twins are given in Table 5. The correlations for IR show a pattern that is typical for many personality variables, with the DZA correlation (0.20) being about half the MZA correlation (0.37). ER shows a much less regular pattern, with the DZA correlation (0.38) being large than the MZA correlation (0.24).

The interclass correlations for spouses based on 40 pairs (including some follow-up pairs not used in this study) for the age and sex corrected scores were 0.32 for IR and 0.13 for ER. These assortative mating coefficients are sufficiently low that, given the modest samples of MZA and DZA twins, we chose not to include assortative mating in our modeling.

The results of model fitting the variance–covariance matrices for the age and sex corrected IR and ER scores are shown in Table 6. Consistent with what one would infer from the intraclass correlations, the

Table 3 Means and standard deviations for the Intrinsic and Extrinsic Measures of Religiosity for MZA and DZA twins

Scale	MZA (n = 70)		DZA n = 74)	
	Mean	SD	Mean	SD
Intrinsic Religiosity (IR)	47.7	15.7	47.8	14.3
Extrinsic Religiosity (ER)	51.6	11.3	54.1	13.5

Table 4 Correlations between the absolute differences in Intrinsic (IR) and Extrinsic (ER) Religiousness scores and measures of Contact and Separation for MZA and DZA twins

Contact measures	MZA (n = 35)		DZA (n = 37)	
	IR	ER	IR	ER
Time Together Prior to Separation (days)	0.09	-0.12	-0.27	-0.33 <sup>a</sup>
Total Contact Time (weeks)	-0.15	-0.11	-0.27	-0.33 <sup>a</sup>
Time Apart (months)	-0.05	-0.04	0.19	0.37 <sup>a</sup>

<sup>a</sup>P<0.05

general model fitted the data much better for both IR and ER than a purely environmental model, even though the latter could not quite be rejected for either variable. The  $\chi^2$  for the general model are small as are the AIC indices indicating a good fit. When the genetic parameter is dropped from the model there is a significant increase in the  $\chi^2$  statistic for both the IR ( $\chi^2 = 8.23$ , (1df),  $P < 0.01$ ) and ER ( $\chi^2 = 6.06$ , 1(df),  $P < 0.01$ ) scales indicating poorer fit. The standardized variance estimates and 95% confidence interval for the two parameters for both variables are given in Table 5.

The correlation between the IR and ER scales and the CPI scales are shown in Table 7. Although a few of the correlations reach statistical significance due to the large sample size, they are all very modest. Because of the large number of correlations we only discuss those significant at  $P > 0.01$ . Self-Control, Femininity/Masculinity, Responsibility and Good Impression correlate 0.21, 0.21, 0.20 and 0.18 with IR. Self-Acceptance, Independence and Dominance have negative correlations of -0.21, -0.20 and -0.19 with ER. There are no significant correlations between either IR or ER and the Vector scales which reflect the core theoretical constructs that purportedly underlie covariation in the scales of the CPI.

Of particular interest is the virtual lack of any significant correlation between the Dickens measures of response style (DSD and DAC) and IR and ER. As noted above, Good Impression, however, has a correlation of 0.18 with IR.

The correlations between the MPQ scales and the IR and ER scales are given in Table 8. There is a negative correlation between Aggression and IR (-0.30) and a positive correlation between Traditionalism and IR (0.37). The higher order Constraint factor scale correlates 0.32 with IR. The largest MPQ

correlate of ER is 0.15 for the Harm Avoidance scale. Neither IR nor ER correlates significantly with any of the response style indices of the MPQ, including the Unlikely Virtues scale which is a measure of social desirable responding, and True Response Inconsistency which is a content free measure of acquiescence.

In order to compare our sample with Altemeyer's, we correlated his measure of RWA (which correlates 0.71 with the MPQ Traditionalism scale) with IR and ER. The correlations (n = 171) were 0.47 and -0.10 for IR and ER, respectively. Since IR correlates 0.47 with RWA and 0.37 with Traditionalism it is clear that facets of IR overlap with the construct of Traditionalism. We also computed the correlation between the MMPI Religious Fundamentalism (REL) scale used in the Walter et al<sup>32</sup> study with the IR and ER scales. The correlations (n = 195) were 0.68 and -0.22.

## Discussion

The alpha reliabilities for IR (0.88) and ER (0.66) are comparable with those found in other adult samples. In Altemeyer's<sup>29</sup> student sample the values were 0.86 and 0.56, and in his parent sample they were 0.89 and 0.72. Gorsuch<sup>26</sup> reports alphas of 0.73 and 0.66 for an adult religious (members of a church) sample (n = 101). The ER scale is clearly less internally consistent than is desirable and could probably be improved with additional item refinement. The correlation of 0.04 between the IR and ER scales shows clear discriminant validity and is reasonably similar to the value of -0.16 reported by Donahue<sup>23</sup> in his meta-analysis of IR and ER correlations for non-religious samples. Note that since all the IR and ER items are stated in a positive fashion, acquiescent response set should result in a positive correlation between the two scales. These results confirm the view that IR and ER assess different dimensions of religiousness in unselected populations and justifies the use of two measures of religiousness rather than one. The slight rewording of the items and the expanded response format used in this study does not appear to have changed the psychometric properties of the scales.

Table 5 Intraclass correlations for MZA and DZA twins and standardized variance estimates for the AE model with 95% confidence intervals for the Intrinsic and Extrinsic Religiousness Scales

Scale	Intraclass correlations		Variance estimates	
	MZA	DZA	V <sub>A</sub>	V <sub>E</sub>
Intrinsic Religiousness (IR)	0.37 (0.04 - 0.63)	0.20 (-0.13 - 0.49)	0.43 (0.15 - 0.64)	0.57 (0.36 - 0.85)
Extrinsic Religiousness (ER)	0.24 (-0.10 - 0.53)	0.38 (0.06 - 0.63)	0.39 (0.09 - 0.61)	0.61 (0.39 - 0.92)

Table 6 Chi-squares and AIC indices derived from fitting two models to the MZA and DZA covariance matrices for the Intrinsic and Extrinsic Religiosity Scales

Scale	Full model (df = 4)			Environmental model (df = 5)		
	$\chi^2$	P	AIC	$\chi^2$	P	AIC
IR	1.00	0.91	-6.99	9.23	0.10	-0.77
ER	4.31	0.37	-3.69	10.37	0.07	0.37

Table 7 Correlations between the Intrinsic Religiosity (IR) and Extrinsic Religiosity (ER) scales and the scales of the California Psychological Inventory (CPI) for the 195 participants who completed both instruments

CPI scales	IR	ER
<b>Folk scales</b>		
Dominance (Do)	-0.07	-0.19 <sup>a</sup>
Capacity for Status (Cs)	-0.02	-0.11
Sociability (Sy)	-0.04	-0.12
Social Presence (Sp)	-0.16	-0.14
Self-Acceptance (Sa)	-0.15	-0.21 <sup>a</sup>
Independence (In)	-0.10	-0.20 <sup>a</sup>
Empathy (Em)	-0.05	-0.16
Responsibility (Re)	0.20 <sup>a</sup>	-0.10
Socialization (So)	0.17	0.06
Self-control (Sc)	0.21 <sup>a</sup>	0.01
Good Impression (Gi)	0.18 <sup>a</sup>	0.08
Communality (Cm)	0.08	-0.18
Well-being (Wb)	0.04	-0.06
Tolerance (To)	0.09	-0.08
Achievement via Conformance (Ac)	0.12	-0.07
Achievement via Independence (Ai)	-0.11	-0.18
Intellectual Efficiency (Ie)	-0.06	-0.22 <sup>a</sup>
Psychological-mindedness (Py)	-0.05	-0.13
Flexibility (Fx)	-0.09	-0.14
Femininity/Masculinity (MF)	0.21 <sup>a</sup>	0.09
<b>Vector scales</b>		
V.1 Internality	0.11	0.13
V.2 Norm-Favoring	0.15	0.12
V.3 Self-Realization	0.03	-0.08
<b>Special Purpose Scales</b>		
Dicken Social Desirability (DSD)	0.13	0.00
Dicken Acquiescence (DAC)	-0.10	-0.02

<sup>a</sup>p>0.01

The correlations between IR and ER and RWA (0.47 and -0.10) replicate Altemeyer's findings very closely, telling us that the scales behave similarly in our adult sample. Traditionalism also correlates much higher with RWA than IR (0.71 vs 0.37), telling us that IR is sufficiently distinct from RWA to justify our treatment of it as a distinctive measure. The correlations between IR and ER and REL (0.67 and -0.22) tell us that IR is measuring something similar to REL, a measure with some construct validity,<sup>32</sup> whilst ER is measuring something else entirely.

The assortative mating coefficients for IR (0.32) is slightly lower than one might expect given that Waller *et al*<sup>32</sup> report assortative mating coefficients of around 0.56 for a variety of measures of religiosity utilizing a somewhat larger group of MISTRA partic-

Table 8 Correlations between the Intrinsic Religiosity (IR) and Extrinsic Religiosity (ER) scales and the scales of the Multidimensional Personality Questionnaire (MPQ) for the 195 participants who completed both instruments

MPQ Scales	IR	ER
<b>Primary Scales</b>		
Well-Being	-0.03	0.01
Social Potency	-0.11	-0.09
Achievement	0.05	-0.13
Social Closeness	0.06	-0.02
Stress Reaction	0.08	0.05
Alienation	-0.04	0.06
Aggression	-0.30 <sup>a</sup>	-0.03
Control	0.17	0.06
Harm Avoidance	0.09	0.14
Traditionalism	0.37 <sup>a</sup>	0.04
Absorption	0.02	0.03
<b>Higher Order Factors</b>		
Positive Emotionality	-0.03	-0.07
Negative Emotionality	-0.01	0.05
Constraint	0.32 <sup>a</sup>	0.10

<sup>a</sup>p>0.01

ipants. The assortative mating coefficient of ER is 0.13. The higher value for IR is consistent with the expectation of assortment for intrinsic religiosity, whereas the lower value for ER is consistent with findings for personality traits. Bouchard *et al*<sup>45</sup> report a mean spousal correlation of 0.22 for the 20 scales of the CPI for 111 spouse pairs in the MISTRA sample. Assortative mating on the MPQ is much less (0.13)<sup>46</sup> if Traditionalism is removed from the list. Additional assortative mating data on IR and ER derived from larger samples would be highly desirable.

The finding of some correlation between contact and separation on ER in the DZA sample is a puzzle because previous analyses of MISTRA data have been unable to document any correlation between contact and twin similarity. Recall also that except for Time Together Prior to Separation these measures are rough approximations because the AUROS was mailed to many of the twins after the assessment at which the contact measures were determined. Contact between the twins does not explain any of the twin similarity on the CPI scales<sup>45</sup> which were all administered at assessment. That the influence of contact is largely restricted to ER for the DZA twins is consistent with the larger DZA intraclass correlation (0.38) for these twins relative to the MZA correlation (0.24). The failure to find later contact effects given the effect of very early contact is, however, curious. We have another way to partially check on the findings of no influence due to later contact. We are currently conducting a ten-year follow-up on MISTRA participants and have AUROS scores on ten additional pairs of MZAs and six additional pairs of DZAs. These twins completed the AUROS only once, at the time of the follow-up.



When these twins are added to the current samples, the MZA and DZA intraclass correlations for IR are 0.35 and 0.17. The comparable correlations for ER are 0.30 and 0.37. Adding twin pairs with additional contact hardly changes the correlations reported in Table 5. Given these results we are inclined to interpret the MZA data as a failure to replicate the DZA results and ascribe the findings to chance. The possibility that ER (and perhaps other similar traits) is easily influenced by contact should be kept in mind. The genetic findings for E should be considered tentative at best.

The similarity of the MZA twins on the MRE scale of the FES (0.32) is larger than found previously (0.18) in an earlier and smaller sample of MISTRA twins.<sup>47</sup> Nevertheless, our analysis of the adopted vs biological family correlations demonstrates that placement on this variable can explain only a trivial amount of the similarity between our twins as MRE is only weakly correlated with religiousness and placement on MRE is only modest. Only environmental variables for which there is both strong selective placement and a strong influence on outcome can explain twin similarity on that outcome. On both counts MRE does not provide an explanation for twin similarity on IR and ER.

Overall, our findings that IR and ER both demonstrate a modest degree of genetic influence confirm previous behavioral genetic findings in the domain of attitudes. The hypothesis that IR would show a higher heritability than ER is not confirmed. Data from a sample of twins reared together could confirm or refute these findings, and gathering such data would clearly be worthwhile.

The only additional family data on IR and ER we have been able to locate is a parent–offspring correlation of 0.39 for IR reported by Altemeyer.<sup>29</sup> He does not report the comparable correlation for E. Since the offspring in his study are still young this correlation could be consistent with a number of hypotheses. Altemeyer argues that ‘the religion of authoritarianism is all of a piece, a self serving belief system, acquired in childhood and strongly reinforced thereafter, which requires no external confirmation and which can probably survive intact in the face of considerable disconfirmation’ (p 222). This is a classic example of interpreting a familial correlation as evidence of environmental causation. A parent–offspring correlation of 0.39 does not refute the hypothesis of a moderate genetic influence nor does it support the hypothesis of complete cultural transmission put forth by the author. The same hypothesis of parental rearing as the determinant of variance in RWA, also put forth by Altemeyer, has been shown to be inconsistent with a behavioral genetic analysis of data from MZ and DZ twins reared together and apart. RWA appears to

have a heritability in the range of 0.50 and 0.64 and modest shared environmental influence.<sup>48</sup> Altemeyer,<sup>31</sup> however, reports a parent–offspring correlation of 0.55 ( $n = 75$ ) for a sample of adoptive parent–offspring pairs, a value that is significantly higher than the biological parent–offspring correlation. It will be interesting to see if this correlation can be replicated in an older sample, as we know that DZ correlations for attitudes drop dramatically once the offspring leave the nest.<sup>18</sup>

The DZA correlations for IR and ER were not less than half the MZA correlations. Thus the hypothesis of non-additivity for religiousness measures suggested by previous research with this sample<sup>32,49</sup> was not confirmed with a different measure of the underlying construct. Non-additivity is a form of interaction, and interactions are notoriously difficult to replicate. A similar failure to replicate non-additive genetic variance for scales of the Multi-dimensional Personality Questionnaire (MPQ) can be seen when the study of Tellegen *et al*<sup>50</sup> is compared with Finkel and McGue.<sup>51</sup>

The failure to find substantive correlations between IR and ER and the scales of the CPI and MPQ is strong evidence for discriminant validity and solidly confirms the unique nature of these measures of religiousness and the need to locate them in the conceptual space of human individual differences. The modest positive correlations between the social maturity cluster of the CPI (Re, So, Sc) and IR are consistent with reports that religiousness is a protective factor against numerous debilitating psychological conditions and should be interpreted as evidence of construct validity, as So and Sc are predictors of delinquency.<sup>52</sup> The correlations between IR and the MPQ scales of Aggression (–0.30) and Traditionalism (0.37) should be interpreted in the same manner. Correlations of this magnitude and in this direction are just what one would expect. The CPI scales used in this study have been shown to be moderately heritable (mean  $h^2 = 0.46$ ) using MZ and DZ twins reared apart and together.<sup>45</sup> The lack of sizeable correlation between the two sets of heritable traits demonstrates that different genetic factors are probably at work in each instance.

The failure of almost all of the response style measures to correlate with IR and ER clearly confirms previous work suggesting that the IR and ER scale scores do not simply reflect invalid response artifacts. Gi was the only exception and the correlation with IR was modest and not replicated by the Unlikely Virtues scale of the MPQ. Chau *et al*<sup>28</sup> also examined personality correlates of the IR and ER scales and concluded that ‘while the I–E scale could be improved upon, it assesses important dimensions of individual differences and major criticisms of the

scale are unsupported or only marginally supported in the data presented herein' (p 400). Our findings and those of Chau *et al* support Donahue's<sup>23</sup> contention that IR is a substantive construct of significant social relevance rather than simply a reflection of conformity, socially desirable responding or acquiescent response style. Our results, however, do not contribute much if anything to the construct validity of ER.

Religiousness is a positive predictor of physical health,<sup>12</sup> a consistent predictor of drug non-involvement in adults<sup>13,53</sup> and adolescents,<sup>14</sup> is related in a complex manner to issues of prejudice and intolerance<sup>23,24,29</sup> as well as mental health<sup>54</sup> and delinquency.<sup>23,55</sup> It is clearly an important construct, deserving of greater attention from social scientists than it has received to date.

## Acknowledgements

This research has been supported by grants from The Seaver Institute, The Pioneer Fund, The University of Minnesota Graduate School, The Koch Charitable Foundation, The Spencer Foundation, The National Science Foundation (BNS-7926654), and the Harcourt Brace Jovanovich Publishing Co.

## References

- 1 Wilson EO. *On Human Nature*. Harvard University Press: Cambridge, 1978.
- 2 Goodenough WH. Evolution of the human capacity for beliefs. *Zygon* 1993; 28: 5–27.
- 3 Maser JD, Gallup GG Jr. Theism as a by product of natural selection. *JReligion* 1990; 70: 515–532.
- 4 Barkow JH, Cosmides L, Tooby J. *The Adapted Mind*. Oxford University Press: Oxford, 1992.
- 5 Buss DM. *Evolutionary Psychology: The New Science of the Mind*. Allyn Bacon: Boston, 1999.
- 6 James W. *Varieties of Religious Experience*. Longman: New York, 1902.
- 7 James W. *The Principles of Psychology*. Holt: New York, 1890.
- 8 Allport GW. *Personality: A Psychological Interpretation*. Holt, Rinehardt & Winston: New York, 1937.
- 9 Ehrlichman H, Brody N. *Personality Psychology: The Science of Individuality*. Prentice Hall: Upper Saddle River, NJ, 1998.
- 10 Mathews G, Deary IJ. *Personality Traits*. Cambridge University Press: Cambridge, 1998.
- 11 Thomas DL, Carver C. Religion and adolescent social competence. In: Gullota TP, Adams GR, Montemayor R (eds). *Developing Social Competence in Adolescence*. Sage: Newbury Park, CA, 1990, pp 195–219.
- 12 Comstock GW, Partridge KB. Church attendance and health. *J Chronic Dis* 1972; 25: 665–672.
- 13 Gorsuch RL. Psychology of religion. In: Rosensweig MR, Porter LW (eds). *Annu Rev Psychol* 1988, pp 201–221.
- 14 Wallace JW, Williams DR. Religion and adolescent health-compromising behavior. In: Schulenberg J, Maggs JL, Hurrellmann K (eds). *Health Risks and Developmental Transitions during Adolescence*. Cambridge University Press: New York, 1997, pp 444–468.
- 15 Meyers SM. An interactive model of religiosity inheritance: The importance of family context. *Am Social Rev* 1996; 61: 858–866.
- 16 Plomin R. *Nature and Nurture*. Brooks/Cole: Pacific Grove, Ca., 1990.
- 17 Cohen D. *Stranger in the Nest*. Wiley: New York, 1999.
- 18 Eaves L, Martin N, Heath A, Schieken R, Meyer J, Silberg J, Neale M, Corey L. Age changes in the causes of individual differences in conservatism. *Behav Genet* 1997; 27: 121–124.
- 19 Wilson EO. Integrated science and the coming century of the environment. *Science* 1998; 279: 2048–2049.
- 20 D'Onofrio BM, Eaves LJ, Murrelle L, Maes HH, Spilka B. Understanding biological and social influences on religious affiliation, attitudes and behaviors: A behavior–genetic perspective. *JPersonality* (in press).
- 21 Lykken DT. Statistical significance in psychological research. *Psychol Bull* 1968; 70: 151–159.
- 22 Williams DR. The measurement of religion in epidemiologic studies. In: Levin JS (ed.). *Religion in Aging and Health: Theoretical Foundations and Methodological Frontiers*. Sage: Newbury Park, CA, 1994, pp 125–148.
- 23 Donahue MJ. Intrinsic and extrinsic religiousness: Review and meta-analysis. *JPers Soc Psychol* 1985; 48: 400–419.
- 24 Allport GW, Ross JM. Personal religious orientation and prejudice. *JPers Soc Psychol* 1967; 5: 432–443.
- 25 Rokeach M. *The Nature of Values*. Free Press: New York, 1973.
- 26 Gorsuch RL, Venable GD. Development of an 'Age Universal' I–E scale. *JSci Study Religion* 1983; 22: 181–187.
- 27 Allport GW, Vernon PE, Lindzey G. *Manual for the Study of Values* (3rd edn). Houghton Mifflin: Boston, 1960.
- 28 Chau LL, Johnson RC, Bowers JK, Darvill TJ, Danko GP. Intrinsic and extrinsic religiosity as related to conscience, adjustment, and altruism. *Pers Individ Diff* 1990; 11: 397–400.
- 29 Altemeyer B. *Enemies of Freedom*. Jossey-Bass: San Francisco, 1988.
- 30 Hutchinson GT, Patock-Peckham JA, Cheong JW, Nagoshi CT. Personality predictors of religious orientation among Protestant, Catholic, and non-religious college students. *Pers Individ Diff* 1998; 24: 145–151.
- 31 Altemeyer R. *The Authoritarian Specter*. Harvard University Press: Cambridge, MA, 1996.
- 32 Waller NG, Kojetin BA, Bouchard TJJr, Lykken DT, Tellegen A. Genetic and environmental influences on religious interests, attitudes, and values: A study of twins reared apart and together. *Psychol Sci* 1990; 1: 1–5.
- 33 Bouchard TJJr, Lykken DT, McGue M, Segal NL, Tellegen A. Sources of human psychological differences: The Minnesota study of twins reared apart. *Science* 1990; 250: 223–228.
- 34 Segal NL. *Entwined Lives: Twins and what they tell us about Human Behavior*. Dutton: New York, 1999.
- 35 Lykken DT, McGue M, Tellegen A. Recruitment bias in twin research: The rule of two thirds reconsidered. *Behav Genet* 1987; 17: 343–362.
- 36 Gough HG. *CPI Manual* (3rd edn). Consulting Psychologists Press: Palo Alto CA, 1996.
- 37 Dahlstrom GW. The growth in acceptance of the MMPI. *Prof Psychol Res Pract* 1992; 23: 345–348.
- 38 Moos RH, Moos BS. *Family Environment Scale: Manual* (2nd edn). Consulting Psychologists Press: Palo Alto, CA, 1986.

- 39 McGue M, Bouchard TJ Jr. Adjustment of twin data for the effects of age and sex. *Behav Genet* 1984; 14: 325–343.
- 40 Taylor HF. *The IQ Game: A Methodological Inquiry into the Heredity Environment Controversy*. Rutgers University Press: New Brunswick, NJ, 1980.
- 41 Farber SL. *Identical Twins Reared Apart: A Reanalysis*. Basic books: New York, 1981.
- 42 Neale MC, Cardon LR (eds). *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic Publishers: Dordrecht, 1992.
- 43 Neale MC. *Mx: Statistical Modeling*. Department of Human Genetics, Box 3 MCV: Richmond, VA, 1995.
- 44 Akaike H. Factor analysis and AIC. *Psychometrics* 1987; 52: 317–332.
- 45 Bouchard TJ Jr, McGue M, Hur Y-M, Horn JM. A genetic and environmental analysis of the California Psychological Inventory using adult twins reared apart and together. *Eur J Pers* 1998; 12: 307–320.
- 46 Lykken DT, Tellegen A. Is human mating adventitious or the result of lawful choice? A twin study of mate selection. *JPers Soc Psychol* 1993; 65: 56–68.
- 47 Hur Y-M, Bouchard TJ Jr. Genetic influence on perceptions of childhood family environment: A reared apart twins study. *Chil Dev* 1995; 66: 330–345.
- 48 McCourt K, Bouchard TJ Jr, Lykken DT, Tellegen A, Keyes M. Authoritarianism revisited: Genetic and environmental influence examined in twins reared apart and together. *Pers Individ Diff* (in press).
- 49 DiLalla DL, Carey G, Gottesman II, Bouchard TJ Jr. Personality indicators of psychopathology via MMPI in twins reared apart. *J Abnorm Psychol* 1996; 105: 491–499.
- 50 Tellegen A, Lykken DT, Bouchard TJ Jr, Wilcox KJ, Segal NL, Rich S. Personality similarity in twins reared apart and together. *JPers Soc Psychol* 1988; 54: 1031–1039.
- 51 Finkel D, McGue M. Sex differences and nonadditivity in heritability of the multidimensional Personality Questionnaire scales. *JPers Soc Psychol* 1997; 72: 929–938.
- 52 Gough HG. Appraisal of social maturity by means of the CPI. *J Abnorm Psychol* 1966; 71: 189–195.
- 53 Gorsuch RL. An interactive, multiple model approach to illicit drug use. In: Lettieri DJ (ed.). *Theories of Drug Use*. National Institute of Drug Abuse: Washington, DC, 1980.
- 54 Bergin AE. Religiosity and mental health: A critical reevaluation and meta-analysis. *Prof Psychol Res Pract* 1983; 14: 170–184.
- 55 Stark R, Kent L, Doyle DP. Religion and delinquency: The ecology of a lost relationship. *J Res Crime Delinquency* 1982; 19: 4–24.