

On the Robustness of Reciprocal Associations Between Personality and Religiosity in a  
German Sample

Richard E. Lucas<sup>1</sup> & Julia Rohrer<sup>2</sup>

<sup>1</sup> Department of Psychology, Michigan State University

<sup>2</sup> Department of Psychology, University of Leipzig

Author Note

Correspondence concerning this article should be addressed to Richard E. Lucas, 316  
Physics Rd., Michigan State University, East Lansing, MI 48823. E-mail: lucasri@msu.edu

## Abstract

Here's an abstract.

*Keywords:* personality, religiosity, cross-lagged panel model

## On the Robustness of Reciprocal Associations Between Personality and Religiosity in a German Sample

A common goal in personality research is to identify robust associations between personality characteristics and consequential behaviors and outcomes. Identifying these associations allows researchers to develop and test hypotheses that inform personality theories. For instance, researchers may study the links between a trait like conscientiousness and an outcome like job achievement. A greater understanding of the processes underlying this association can inform theories about conscientiousness. In addition, this research could provide practical guidance for those seeking to improve achievement levels. Moreover, a consideration of the reverse causal direction—understanding whether achievement experiences impact trait levels—can inform theories of personality development and change. Thus, studies that examine the processes underlying such links have great potential further the understanding of how personality characteristics shape peoples lives, and how life experiences shape personality.

Recently, Entringer, Gebauer, and Kroeger (2023) conducted such an examination, investigating the links between the Big Five personality traits and religiosity in a very large German sample. This research was motivated both by prior theories meant to explain how personality can shape religiosity and by theories that posit that religiosity can affect personality. For instance, a niche-picking perspective suggests that people who have personality traits that are consistent with the behaviors that are typically exhibited in religious contexts should gravitate towards these religious contexts. In addition, the Sociocultural Norm Perspective (Eck & Gebauer, 2022) posits that personality traits like agreeableness, conscientiousness, and (low) openness to experience produce normative behaviors; when religiosity is normative in a culture, then these traits should cause greater religiosity. In contrast, complementary theories suggest that religiosity itself can impact these same traits, as religious contexts also promote or even enforce behaviors and views that

are consistent with traits like agreeableness and conscientiousness.

To test these ideas, Entringer et al. (2023) relied on a widely used approach for examining reciprocal causal effects in panel data: the cross-lagged panel model (CLPM, Heise (1970)). In the CLPM, each variable (in this case, personality and religiosity) at each occasion is predicted from the same variables assessed at prior waves. With some assumptions, lagged associations from one variable to the other (e.g., from Time 1 personality to Time 2 religiosity) can be interpreted as causal effects. Entringer et al. (2023) found that agreeableness, openness, and conscientiousness prospectively predicted changes in religiosity (at least in some contexts), whereas religiosity predicted changes in agreeableness and openness (again, in some contexts). Moreover, the religiosity of the region in which respondents lived moderated the links between some personality traits and religion.

Entringer et al.'s study had a number of desirable features that make it especially well-suited to examining questions about reciprocal associations between personality and religiosity. First, the authors used a very large panel study with four waves of assessment over a period of twelve years. These features should contribute to the robustness of the results. Moreover, the authors used a sophisticated latent-variable version of the CLPM that accounts for measurement error, which can help reduce the likelihood of spurious lagged associations (Lucas, 2023). In addition, the authors examined these associations separately in different German federal states that varied in overall religiosity, which allowed them to examine theoretically relevant contextual moderators of these associations. Finally, the authors conducted many robustness checks to support their primary findings.

### **Concerns About Robustness**

Despite these strengths, however, there are reasons to be concerned about the robustness of the support for reciprocal causal effects between personality and religiosity. The first issue concerns the size of the effects that Entringer et al. (2023) found. The

authors foreshadow this concern early in the paper, noting that lagged effects are typically much smaller than cross-sectional effects and that prior cross-sectional correlations should provide an upper bound on the size of any lagged associations. Because the correlations between personality and religiosity tend to be small (e.g., around .19 in their review), the lagged effects should be even smaller. Indeed, the observed lagged effects in their study were quite small, with maximum standardized regression coefficients of .039. The estimated moderator effects they found were similarly small in size.

Although we agree that small effects can sometimes be important, such effects provide challenges for interpretation. First, just because such effects can be important in some contexts does not mean that they are always important; justification for why a particular small effect is important is needed. Moreover, one common defense of the importance of small effects is that these effects accumulate over time. However, this should mean that aggregated between-person correlations should themselves be reasonably large in size. Entringer et al. (2023) did not report zero-order correlations either within waves or after aggregating personality and religiosity across the 12-year period. Perhaps more importantly, very small effect sizes can be problematic because subtle model misspecification or residual confounding can lead to small effects. This makes it difficult to distinguish true effects from model misspecification or confounding when effects are small. Additional concerns about model misspecification and confounding could easily lead to the size of effects found in this study.

For instance, although Entringer et al.'s decision to model latent personality factors when examining reciprocal associations has the desirable feature of removing measurement error, it also comes with a cost in terms of model complexity. Although researchers might hope that the items of their measures load cleanly on the factor to which the item belongs (and not to any other), this is not always the case in practice. In such cases, allowing for secondary loadings may be necessary to improve model fit, though questions can remain about whether such post hoc modifications capitalize on chance. If decisions about the

measurement model affect the structural features of the model, then concerns about the robustness of estimates (especially estimates that are very small in size) can be raised. In the current study, we compare the latent-variable model with a complex measurement-model specification from Entringer et al. (2023) to a simpler observed-variable model that includes only the observed mean scores for each personality trait measure.

A second concern is that Entringer et al. (2023) chose to model all five traits simultaneously when predicting changes in religiosity. Although the Big Five traits are hypothesized to be relatively independent, in practice they are not. Thus, when modeling all traits simultaneously, estimated paths from personality to religiosity reflect associations that persist after controlling for all other personality traits. The decision to control for correlated variables comes with interpretational challenges, however, as the association can only be interpreted as an association between religiosity and the variance that is not shared with other traits (Lynam, Hoyle, & Newman, 2006). The conceptual connections between this residualized variable and the hypothetical construct it is meant to assess are not always clear. Although Entringer et al. (2023) justified their decision to simultaneously model all five traits by noting that this decision is consistent with prior research, a more substantive justification would be preferable. Because of the interpretational challenges, our preference is to interpret unadjusted associations; but at the very least, robustness across modeling choices is important to consider.

The final concern is potentially the most consequential. When examining reciprocal effects, Entringer et al. (2023) relied solely on the traditional CLPM with a single lagged association with the prior wave. However, Hamaker, Kuiper, and Grasman (2015) showed that the CLPM results in biased lagged associations when stable-trait variance exists in the measures being modeled. Recently, Lucas (2023) used simulations to show that this problem is quite severe; spurious lagged associations can be found as often as 100% of the time in realistic scenarios (e.g., when there is just a moderate amount of stable-trait variance, when

sample sizes are moderate to large, and when multiple waves of assessment are included)<sup>1</sup>. The size of the bias found in these simulations is high relative to the size of effects reported by Entringer et al. (2023). Notably, two strengths of Entringer et al.’s study (the very large sample size and the use of a four-wave design) increase the likelihood of finding spurious lagged effects.

## The Present Study

Although Entringer et al. (2023) took many steps to ensure the robustness of their results, additional concerns can be raised about effect sizes, model complexity, the decision to simultaneously model all five personality traits, and—most importantly—the decision to use a traditional CLPM instead of a model that accounts for additional sources of stability. The goal of the current analysis is to test the robustness of these results to alternative specifications.

We first simply examine the correlations between religiosity and each of the Big Five traits within each wave and aggregated across all waves. This provides a simple index of effect size that helps establish whether any lagged causal effects accumulate over time. Next, we test a series of models that examine the robustness of the results reported in Entringer et al. (2023). Specifically, after first replicating their results, we then test various combinations of three modifications. We compare models where the Big Five traits are modeled as latent-traits (using the same measurement model in the original paper, including secondary loadings) versus when they are modeled as observed variables. Next, we compare models

---

<sup>1</sup> Recently, Orth, Clark, Donnellan, and Robins (2021) defended the CLPM against concerns raised by Hamaker et al. (2015) and others. Their defense focused on the interpretation of the between-person and within-person associations that are modeled in the alternatives to the CLPM. Specifically, they argued that the CLPM can test “between-person prospective effects,” while the alternatives to the CLPM do not. As explained in Lucas (2023), we disagree that the between-person prospective effects that Orth et al. hope to assess are clearly defined, and we disagree that the CLPM tests them (see Lucas, 2023 for a discussion). More importantly, this defense ignores the fact that a critical assumption of the CLPM is that no additional source of stability in the outcome measures exists beyond those that are included in the model (i.e., the autoregressive effect reflected in the stability of a variable over time and the lagged effect of one variable on the other at a later time). If stable-trait variance exists, then this assumption is violated, invalidating the interpretation of the lagged paths as causal effects (Heise, 1970).

where the Big Five traits are entered simultaneously as predictors to those where separate models are run for each trait individually. Finally, we compare the CLPM to the random-intercept cross-lagged panel model (RI-CLPM, Hamaker et al., 2015), which includes a random intercept to account for stable-trait variance. Combining each pair of comparisons results in eight separate models, with results for each of the Big Five traits. In addition, Entringer et al. (2023) examined the moderating contextual effect of state-level religiosity. We also test these moderating effects in each of the eight models for each of the Big Five traits. The results will then be compared for robustness.

## Methods

This paper uses data from the German Socio-Economic Panel (SOEP) study, which assessed the Big Five personality traits and religiosity four times, at four-year intervals. The inclusion of four waves of assessment allows for the use of both the CLPM and the more complex RI-CLPM. Entringer et al. (2023) provided detailed code for their analyses, however, they did not provide code to extract and clean data from the raw data files. As a test of the reproducibility of the analyses from the description provided in the text, we developed our own code for extracting and cleaning data. Despite our best attempts, the exact number of participants included in the final samples, and the precise estimates from the original model could not be perfectly reproduced. However, both the final sample sizes and results are quite close to the original, and all substantive conclusions from the original paper were supported. Thus we proceeded with our robustness checks using the sample we extracted. Our full code for extracting, cleaning, and analyzing these data is available at [https://github.com/rucas11/gsoep\\_religion](https://github.com/rucas11/gsoep_religion)).

## Participants

The SOEP is a long-running panel study



## Results

We first examine the size of the correlations between each Big Five personality trait and religiosity. To allow any short-term associations to accumulate over time, we focus on the correlations between scores aggregated across all available assessments. Specifically, we computed each person's average score for each personality trait across all available waves and the average religiosity score across all waves. We then examined the correlations among these variables (a) in each state, (b) by pooling the within-state associations, and (c) by examining the raw correlation among all participants, ignoring residency. These correlations are reported in table 1.

As can be seen in this table, even after aggregating over a 12-year period, all correlations are quite small. The largest correlation for the full sample is between religiosity and agreeableness, which is just 0.10. The largest correlation among the 80 tested (including those from the full sample and pooled within-state associations) was just 0.14. Thus, even when allowing an effect to accumulate over a 12-year period, correlations between personality and religiosity are quite small. We next turn to models that examine the reciprocal associations between personality and religiosity.

### Full Sample Analyses

In their original paper, Entringer et al. (2023) focused on results from each federal state individually and on the meta-analytic averages of these state-level associations. This allowed them to examine whether results varied across states and specifically to test whether state-level religion moderated associations between religiosity and personality. Although we also use this approach for some of our analyses, we first present model comparisons in the full sample, ignoring state of residence. We do this for three reasons. First, as we will show, the estimates from the full sample mostly replicate the meta-analytic averages across the 14 analyzed states, and focusing on one sample instead of 14 (or 16) allows for clearer model comparisons. Second (and most importantly), in the original analyses, two states were

Table 1

*Within-state correlations between each personality trait and religiosity.*

State	Correlation with Religiosity					Religiosity	N
	Agr	Con	Ext	Neu	Opn		
Schlewsig-Holstein	0.10	0.13	0.01	-0.01	0.05	1.59	1605
Hamburg	0.08	-0.01	-0.04	-0.02	0.02	1.52	677
Lower Saxony	0.11	0.08	0.00	0.00	0.04	1.73	4459
Bremen	0.07	-0.01	0.05	0.01	0.04	1.65	336
North Rhone-Westphalia	0.13	0.09	-0.01	0.00	0.02	1.78	9968
Hesse	0.14	0.10	0.01	-0.01	0.03	1.80	3300
Rhineland-Palatinate	0.08	0.06	-0.03	0.00	0.01	1.83	2081
Baden-Wuerttemberg	0.10	0.05	-0.02	0.01	-0.01	1.87	5672
Bavaria	0.07	0.06	-0.05	0.00	-0.04	1.96	7248
Saarland	0.09	0.14	0.04	0.02	0.07	1.77	428
Berlin	0.10	0.08	0.06	-0.05	0.08	1.45	1601
Brandenburg	0.08	-0.01	0.01	-0.02	0.11	1.34	1738
Mecklenburg-West Pomerania	0.09	0.04	-0.03	-0.02	0.06	1.28	967
Saxony	0.04	0.00	-0.01	0.01	0.10	1.40	2848
Saxony-Anhalt	0.11	0.07	0.00	0.00	0.14	1.33	1637
Thuringia	0.07	0.04	-0.02	0.03	0.11	1.46	1751
Pooled Within	0.10	0.07	-0.01	0.00	0.03		46316
Raw Correlation	0.10	0.06	-0.01	-0.01	0.03		46316

*Note.* Agr = Agreeableness; Con = Conscientiousness; Ext = Extraversion; Neu = Neuroticism; Opn = Openness. Mean religiosity and sample sizes are presented in the rightmost columns.

dropped because of small sample sizes that resulted in problems with model convergence. When testing additional models, the specific states in which estimation problems emerge differ across models, which means that any differences that result could be due to differences in the models or differences in which states provide estimates. Finally, by ignoring the state-level structure, no participants needed to be excluded due to residence in states for which sample sizes were small. Again, we note that we eventually do discuss results in each state individually in the next section.

The models we tested compare estimates across three dichotomous factors: (1) The original CLPM versus the potentially more appropriate RI-CLPM, (2) Models that include latent personality traits (with a complex measurement model) versus models that include only a single observed mean score for each trait, and (c) models that include all five traits simultaneously (and hence control for correlations among traits when predicting religiosity) versus models that include just one trait per model. The combination of these factors results in eight sets of estimates for each personality trait. Fit indices for all models are presented in Table 2. Consistent with Entringer et al. (2023), we report Chi-square with degrees of freedom and p-value, along with CFI, RMSEA, and SRMR.

Evaluating the fit of these models is challenging, as the chi-square is sensitive to sample size and subtle misfit can lead to significant values with large sample sizes, and there are no unambiguous cutoffs for well-fitting models using the other indices. Recommended cutoffs for the CFI are typically either .90 or .95, and the recommend cutoff for the RMSEA and SRMR is typically .05. The CLPM is nested within the RI-CLPM, so the difference in fit between these pairs of models can be tested explicitly, though the large sample sizes make statistical significance testing overly sensitive. Other pairs of models are not nested and cannot be directly compared.

Consistent with the original paper, our focus is on the cross-lagged paths from each trait to religiosity and from religiosity to each trait. Figure 1 presents the estimated cross-lagged path from each personality trait to religiosity for each of the eight tested models. Results from the CLPM are presented in black lines, whereas results for the RI-CLPM are presented as grey lines. Results from the latent-variable models are included as solid lines whereas results for the observed-variable models are presented as dashed lines. Finally results for models that include all five traits simultaneously are labeled with triangles, whereas those that model each trait individually are labeled with circles. Bars represent 95% confidence intervals for the standardized parameter estimates.

Table 2

*Fit indices for full sample models.*

Model	Type	Variables	Trait	ChiSq	df	p-value	CFI	RMSEA	SRMR
CLPM	Latent	All		44,390.09	1,835	0.000	0.90	0.02	0.05
RICLPM	Latent	All		34,826.20	1,814	0.000	0.92	0.02	0.04
CLPM	Observed	All		11,781.85	210	0.000	0.92	0.03	0.07
RICLPM	Observed	All		854.77	189	0.000	1.00	0.01	0.02
CLPM	Latent	Single	agr	2,876.49	90	0.000	0.96	0.03	0.04
			cns	3,228.93	90	0.000	0.97	0.03	0.04
			ext	2,931.83	90	0.000	0.98	0.03	0.04
			neu	2,456.88	90	0.000	0.97	0.02	0.03
			opn	3,129.27	90	0.000	0.97	0.03	0.04
RICLPM	Latent	Single	agr	709.53	87	0.000	0.99	0.01	0.02
			cns	942.10	87	0.000	0.99	0.01	0.03
			ext	668.99	87	0.000	0.99	0.01	0.02
			neu	305.40	87	0.000	1.00	0.01	0.01
			opn	938.72	87	0.000	0.99	0.01	0.03
CLPM	Observed	Single	agr	3,606.76	22	0.000	0.93	0.06	0.08
			cns	3,478.70	22	0.000	0.93	0.06	0.08
			ext	4,025.79	22	0.000	0.93	0.06	0.08
			neu	3,706.82	22	0.000	0.93	0.06	0.08
			opn	3,704.36	22	0.000	0.93	0.06	0.08
RICLPM	Observed	Single	agr	80.42	19	0.000	1.00	0.01	0.01
			cns	138.37	19	0.000	1.00	0.01	0.01
			ext	136.38	19	0.000	1.00	0.01	0.01
			neu	135.11	19	0.000	1.00	0.01	0.02
			opn	110.95	19	0.000	1.00	0.01	0.01

*Note.* Agr = Agreeableness; Con = Conscientiousness; Ext = Extraversion; Neu = Neuroticism; Opn = Openness.

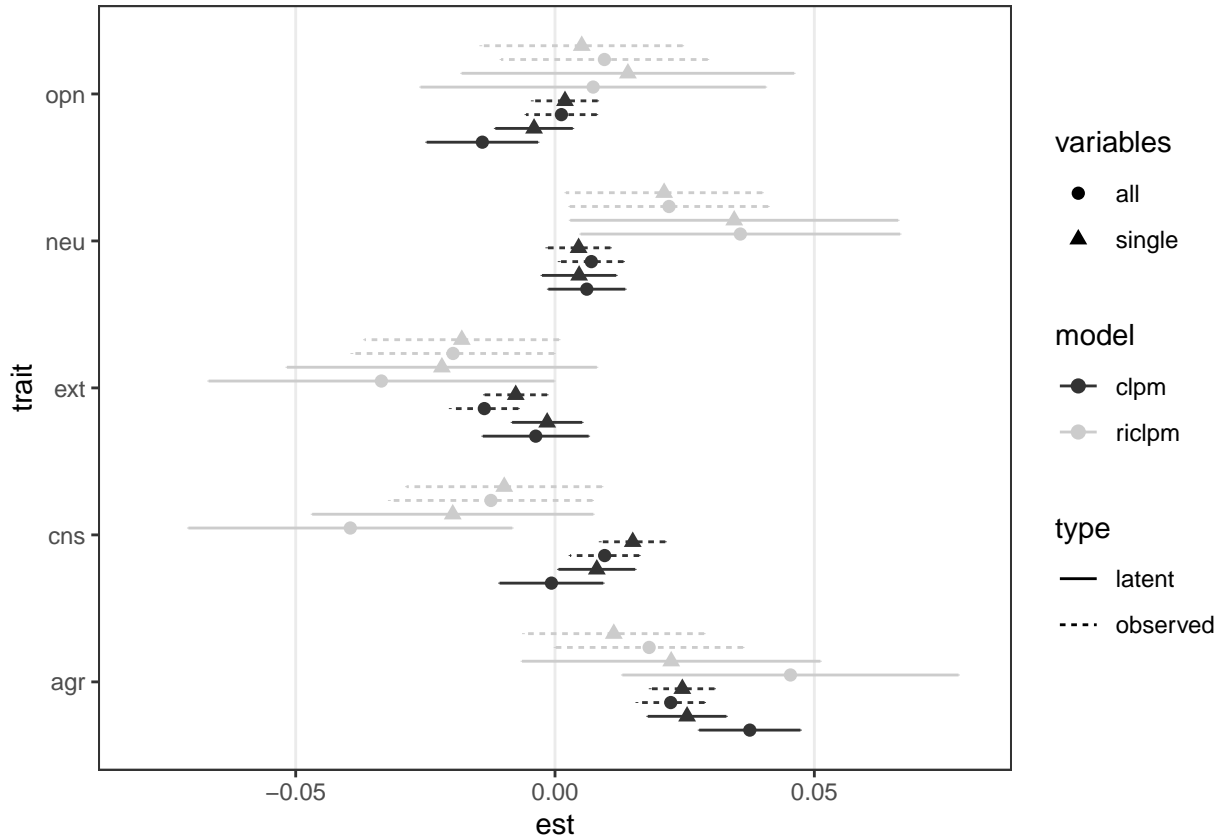


Figure 1. Estimated Lagged Effects of Personality on Religiosity

Before describing the results, it is important to acknowledge that the confidence intervals for estimates from the RI-CLPM models are generally considerably larger than those for the CLPM. This pattern is not unique to these data or these model specifications. The RI-CLPM is a multilevel model that separates between-person associations from within-person associations, and estimates of the within-person parts of these models often have less bias at the cost of reduced efficiency (see Allison, 2009, for an explanation). Because of this, we highlight both the significance of the effect and the parameter estimates when comparing results across models.

The figure shows that conclusions about the lagged effects of personality on religiosity would differ depending on which model was used. Indeed, there is no effect that emerges consistently across all model specifications. For instance, the largest effect from Entringer et

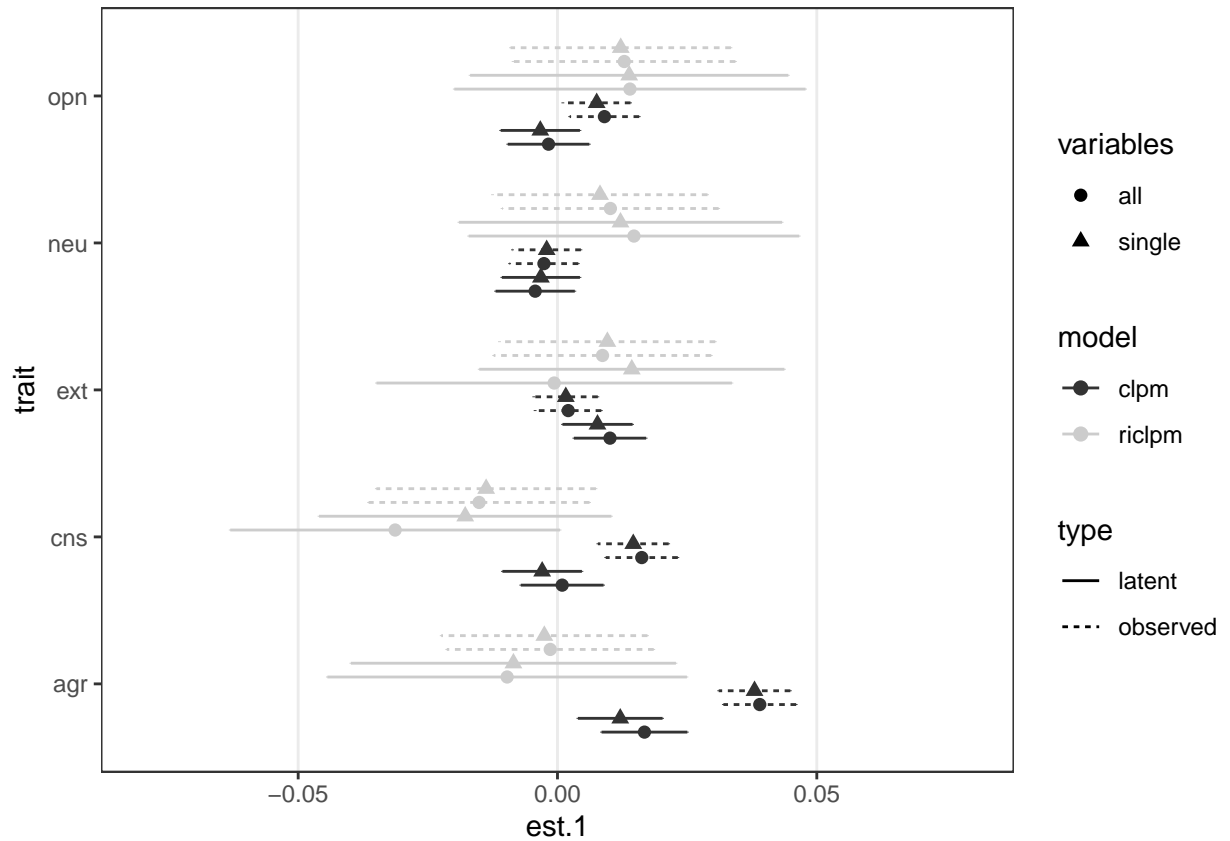
al. (2023) was the lagged association between agreeableness and religiosity, which had an average standardized effect of .039. The comparable model from Figure 1 is the CLPM with latent traits and all traits modeled simultaneously, which resulted in an almost identical average standardized effect of 0.038. This effect was actually slightly larger when the RI-CLPM was used, when latent variables were modeled, and when all traits were included simultaneously. However, estimated effects were smaller (frequently about half the size) and sometimes nonsignificant in the other model specifications. This association between agreeableness and religiosity was the most robust of the effects we examined, and even it varied across model specifications.

Importantly, Figure 1 shows that our alternative models do not always result in reduced effect sizes relative to what was reported in Entringer et al. (2023). For instance, in the original study, the lagged effect of neuroticism on religiosity was a nonsignificant .011. In our reanalysis, we again found similar average estimate of 0.006 with the same model specification in the full sample. However, the size and significance of the effect varied considerably across specifications. Most notably, the RI-CLPM with latent variables and all variables entered simultaneously resulted in an estimated lagged effect that was approximately the same size as the lagged effect of agreeableness (the largest effect found in the original study and a primary finding that was highlighted in the discussion). It is sometimes claimed that using the RI-CLPM underestimates causal effects (e.g., Asendorpf, 2021), but this is not correct. Lucas (2023) showed that if stable-trait variance exists in the measures, results from the CLPM can underestimate the true lagged effects. Neuroticism received little attention in the original report, even though the lagged path from neuroticism to religiosity is one of the largest effects found when the arguably more appropriate RI-CLPM is used.

Finally, Figure 1 shows that effects can even reverse direction depending on the chosen model specification. Most notably, although there was no significant effect of

conscientiousness in the original study (a finding replicated in the full sample using the same model specification), this effect became significantly negative in one version of the RI-CLPM and significantly positive in the other three specifications of the CLPM. These examples show that conclusions about the lagged associations between personality traits and religiosity depend on the precise model specification. Indeed, conclusions about all five traits depend on which model specification is chosen (if the significance of the effect is used to guide interpretation), with the sign of the effect even changing for one of the five traits.

Entringer et al. (2023) found only one significant lagged association from religiosity to personality: the path from religiosity to agreeableness, which was estimated to be a small but significant .011. This effect was again replicated in our full-sample analysis (average standardized estimate = 0.017). However, the size and significance of this effect again varied considerably; two specifications of the CLPM resulted in considerably larger effects (with effect sizes comparable to the largest effects of personality on religion) and nonsignificantly negative associations in all specifications of the RI-CLPM. Again, when relying on the significance of the effect to guide conclusions, these would differ for three of the five traits assessed, and for two of these, the differences in the size of the estimates were considerable.



## Disclosures

### Author Contributions

Richard E. Lucas conceptualized the study and wrote the initial analysis code and the first draft of the paper.

Julia Rohrer wrote additional code and ran all analyses, contributed additional ideas for analyses, and contributed to writing and editing the text.

### Conflicts of Interest

The author declares that there were no conflicts of interest with respect to the authorship or the publication of this article.



**Prior Versions**

A preprint of this paper was posted on the PsyArXiv preprint server: .

### References

- Allison, P. D. (2009). *Fixed effects regression models*. Sage.
- Asendorpf, J. B. (2021). Modeling developmental processes. In *The Handbook of Personality Dynamics and Processes* (pp. 815–835). Elsevier.  
<https://doi.org/10.1016/B978-0-12-813995-0.00031-5>
- Entringer, T. M., Gebauer, J. E., & Kroeger, H. (2023). Big Five Personality and Religiosity: Bidirectional Cross-Lagged Effects and their Moderation by Culture. *Journal of Personality*, *n/a*(*n/a*). <https://doi.org/10.1111/jopy.12770>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. <https://doi.org/f67cvh>
- Heise, D. R. (1970). Causal Inference from Panel Data. *Sociological Methodology*, *2*, 3–27.  
<https://doi.org/10.2307/270780>
- Lucas, R. E. (2023). Why the Cross-Lagged Panel Model Is Almost Never the Right Choice. *Advances in Methods and Practices in Psychological Science*.
- Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The Perils of Partialling: Cautionary Tales from Aggression and Psychopathy. *Assessment*, *13*(3), 328–341.  
<https://doi.org/10.1177/1073191106290562>
- Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*(4), 1013–1034. <https://doi.org/gg7zfw>