RELIGION, BRAIN & BEHAVIOR
2023, VOL. 13, NO. 3, 284–363
https://doi.org/10.1080/2153599X.2022.2070251

Routledge
Taylor & Francis Group

COMMENTARIES

Being specific about generalisability

Quentin Douglas Atkinson [a,b], Scott Claessens [a], Kyle Fischer [a], Guy Lavender Forsyth [a], Thanos Kyritsis [a], Kristina Wiebels [a], and David Moreau [a]

[a]School of Psychology, University of Auckland, Auckland, New Zealand; [b]Max Planck Institute for the Science of Human History, Jena, Germany

In the behavioural sciences, big claims about the world are often supported with few observational or experimental findings from particular sample populations. This gap between headlines and their corresponding empirical findings is known as the problem of generalisability (Yarkoni, 2020). But generalisability is not unidimensional: there are many ways in which a finding might generalise. Do findings from one place and time generalise to other places and times (Henrich et al., 2010)? Do particular operationalisations of broad concepts, such as "religiosity" and "wellbeing" (Abdel-Khalek, 2019), generalise to other possible operationalisations? Do findings generalise from one statistical analysis to alternative analytic choices? Rather than asking "does this finding generalise?", researchers must be specific and ask the more difficult question "*in what ways* does this finding generalise?"

The Many Analysts Religion Project (MARP; Hoogeveen et al., 2002) takes what might be called a "scattergun" approach to generalisability across a wide variety of domains (see Figure 1, top panel). Each research team must make decisions about whether or how to account for population structure in their data (comprising 24 nations and more than 19 different religious traditions), how to operationalise variables (across nine items measuring "religiosity" and 18 items measuring "wellbeing", including four wellbeing subdomains), and what analytical techniques and modelling approaches to use to statistically test the hypotheses of interest. By integrating findings from more than 100 independent teams, MARP provides insight into the generalisability of findings across these researcher degrees of freedom (Chambers, 2019).

MARP's freedom for researchers also enables individual teams to deal with the issue of generalisability in their own ways (see Figure 1, bottom panel). Addressing generalisability *within* any analysis is critical, not only because most research projects will not include many analysts, but because generalisability can be explicitly modelled and tested in a common analytical framework. Our two teams independently chose different analytic strategies, yet in separate ways we both gravitated towards testing generalisability.

One team (Team 005, osf.io/352ma) elected to use expansive multilevel modelling to explore the generalisability of religiosity's effects on wellbeing. The multilevel approach provides a statistical framework for exploring sources of variation in the data simultaneously (McElreath, 2020). This strategy drew directly from Yarkoni's (2020) argument that the uncertainty surrounding particular operationalisations should be explicitly included within statistical models. In particular, this team modelled wellbeing as an ordinal outcome variable and included random intercepts and slopes for religiosity that varied across wellbeing items and subscales, religious denominations, and nations, controlling for non-independence of nations due to shared cultural ancestry (Bromham et al., 2018; see Claessens et al., 2021 for full writeup).

Our other team (Team 080, osf.io/z92gr) chose a complementary approach based on model cross-validation, allowing the exploration of generalisability across different partitions folds of

CONTACT Thanos Kyritsis t.kyritsis@auckland.ac.nz

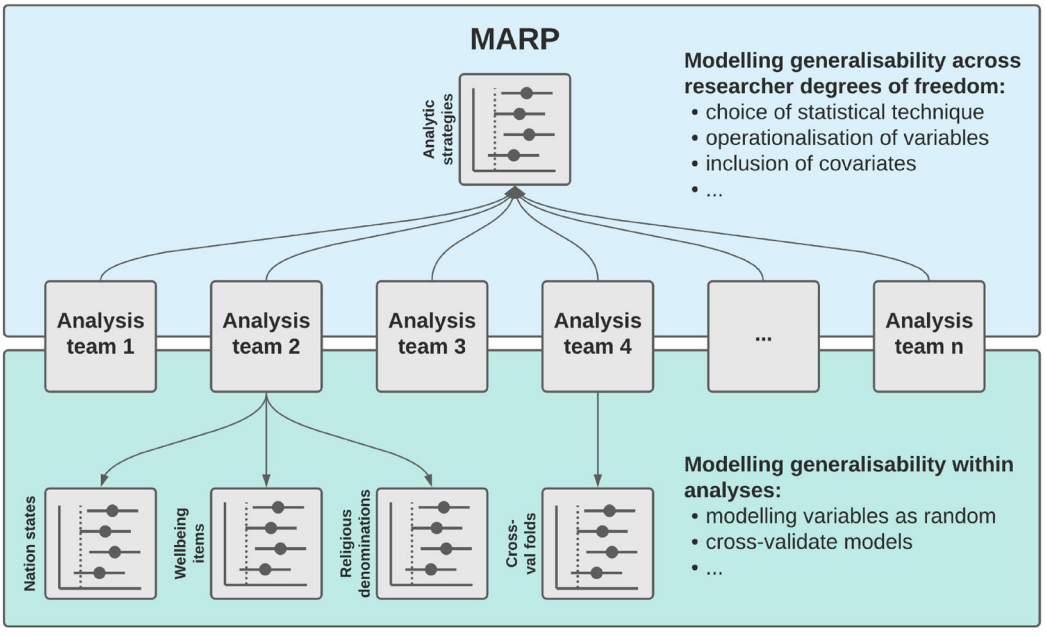

Figure 1. Different kinds of generalisability. The Many Analysts Religion Project allows researchers to explore the generalisability of the relationship between religiosity and wellbeing *across* researcher degrees of freedom, such as different statistical techniques and operationalisation of variables. Generalisability can also be addressed *within* analyses, for example by modelling certain variables as random or by cross-validating models on left-out data. Point ranges represent regression coefficients and 95% credible intervals across different clusters for mock data.

the data (de Rooij & Weeda, 2020). While a particular model of the effects of religiosity on wellbeing may be suited to one dataset, it may not be equally suited to explaining the relationship between religiosity and wellbeing in out-of-sample data. By dividing the dataset into segments, however, models validated on one part of the data can then be tested on the other (Vehtari et al., 2017) in an iterative process that ensures models are not unduly influenced by idiosyncratic properties of the dataset at hand. Empirically testing model fit this way provides reassurance of reliable out-of-sample prediction and thus generalisability across datasets.

The approaches of our two teams are therefore not mutually exclusive, and both are compatible with the overarching goal of generalisability. The multilevel approach favoured by our first team illustrated the scope and limits of religiosity's effect on wellbeing; for example, we found that religiosity predicted most individual wellbeing items, but this effect did not generalise to mobility, physical pain, or medical dependence. The approach preferred by our second team revealed that the inclusion of age, socioeconomic status, and education as covariates was important to maximise the predictive validity of empirical claims.

The bringing together of many researchers in MARP is a great opportunity to explore generalisability in the context of religion and wellbeing. For both the science that builds on these findings and their real-world application, we must be specific about generalisability and ask "*in what ways* do these findings generalise?" The methods we implemented are easily accessible and begin to answer this question. We hope that future research can leverage approaches like these to address the multiple facets of generalisability that exist, not only for the study of religiosity and wellbeing, but across all fields of study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Quentin Douglas Atkinson http://orcid.org/0000-0002-8499-7535
Scott Claessens http://orcid.org/0000-0002-3562-6981
Kyle Fischer http://orcid.org/0000-0001-7204-1125
Guy Lavender Forsyth http://orcid.org/0000-0003-3077-7112
Thanos Kyritsis http://orcid.org/0000-0003-2790-1813
Kristina Wiebels http://orcid.org/0000-0002-5360-5965
David Moreau http://orcid.org/0000-0002-1957-1941

References

Abdel-Khalek, A. M. (2019). Religiosity and well-being. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopaedia of personality and individual differences*. Springer Nature. https://doi.org/10.1007/978-3-319-28099-8_2335-1.

Bromham, L., Hua, X., Cardillo, M., Schneemann, H., & Greenhill, S. J. (2018). Parasites and politics: Why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society Open Science*, *5*(8), 181100. https://doi.org/10.1098/rsos.181100

Chambers, C. (2019). *The seven deadly sins of psychology*. Princeton University Press.

Claessens, S., Fischer, K., Lavender Forsyth, G., Kyritsis, T., & Atkinson, Q. D. (2021). Religion, cultural norms, and wellbeing: A bayesian multilevel analysis [Computer Software]. https://github.com/ScottClaessens/religionWellbeing.

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33* (2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., Altay, S., Alzahawi, S., Amir, Y., Anthony, F.-V., Appiah, O. K., Atkinson, Q. D., Baimel, A., Balkaya-Ince, M., Balsamo, M., Banker, S., Bartoš, F., Becerra, M., Beffara, B., Beitner, J., Bendixen, T.,…Wagenmakers, E.-J. (2022). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior*. https://doi.org/10.1080/2153599X.2022.2070255

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and stan (2nd ed.)*. Chapman and Hall.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, *45*(E1), 1–37. https://doi.org/10.1017/S0140525X20001685

Advantages of using multilevel modeling approaches for the many analysts religion project

Merve Balkaya-Ince and Sarah A. Schnitker

Department of Psychology and Neuroscience, Baylor University, Waco, TX, USA

Cross-cultural research involves the systematic study of psychological processes across different cultures to understand similarities and differences in human behavior as they are influenced by cultural context (Matsumoto & Juang, 2017). Cross-cultural research offers key advantages for informing theories on individual and cultural sources of variation in human behavior because of

CONTACT Merve Balkaya-Ince merve_balkayaince@baylor.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/2153599X.2022.2070264.