Religion, Brain & Behavior

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rrbb20

Resolving religious debates through a multiverse approach

Angelos-Miltiadis Krypotos, Richard A. Klein & Jonathan Jong

To cite this article: Angelos-Miltiadis Krypotos, Richard A. Klein & Jonathan Jong (2023) Resolving religious debates through a multiverse approach, Religion, Brain & Behavior, 13:3, 318-320, DOI: 10.1080/2153599X.2022.2070261

To link to this article: https://doi.org/10.1080/2153599X.2022.2070261

Published online: 06 Jul 2022.

References

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys*, *31*(1), 326–342. https://doi.org/10.1111/joes.12139

European Social Survey Cumulative File, ESS 1-9. (2020). Data file edition 1.0. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC. https://doi.org/10.21338/NSD-ESS-CUMULATIVE.

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/10.1037/a0038889. PMID: 25822208.

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

Johnson, T. M., & Grim, B. J. (Eds.), (2021). *World Religion database*. Brill.

Kim-Prieto, C., & Miller, L. (2018). Intersection of religion and subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. DEF Publishers. doi:nobascholar.com

Maoz, Z., & Henderson, E. A. (2013). The world religion dataset, 1945–2010: Logic, estimates, and trends. *International Interactions*, *39*(3), 265–291. https://doi.org/10.1080/03050629.2013.782306

Witter, R., Stock, W., Okun, M., & Haring, M. (1985). Religion and subjective well-being in adulthood: A quantitative synthesis. *Review of Religious Research*, *26*(4), 332–342. https://doi.org/10.2307/3511048

Resolving religious debates through a multiverse approach

Angelos-Miltiadis Krypotos[a], Richard A. Klein[b], and Jonathan Jong[c]

[a]Clinical Psychology, Utrecht University and KU Leuven, Utrecht, the Netherlands; [b]Department of Methodology and Statistics, Tilburg University, the Netherlands; [c]Institute for Cognitive and Evolutionary Anthropology, Oxford University, Oxford, UK

In recent years, several many-analysts projects have been published, projects in which synthetic or real data are analyzed independently by different experts (e.g., Aczel et al., 2021; Schweinsberg et al., 2021; Silberzahn et al., 2018; Silberzahn & Uhlmann, 2015). Such projects are particularly useful in investigating how analytic variability may influence the strength and direction of results—when inferential statistics are used (e.g., Schweinsberg et al., 2021)—or the estimation of individual model parameters—when computational models are pitted against each other (e.g., Boehm et al., 2018; Dutilh et al., 2019).

The present project (Hoogeveen et al., 2022) tasked multiple teams with testing two specific hypotheses using the same dataset: first, testing the relation between religiosity and well-being; second, testing whether this (potential) relationship depends on the cultural norms of religion. Once again, a key finding is that variability in data analyses leads to different results, with results sometimes even differing in their direction (e.g., on question 2 both negative and positive effects are reported). Importantly, the different results do not necessarily mean that any of the expert teams followed a wrong approach, and we are confident that all included analysts can provide principled argumentations as to why they followed their chosen statistical approach.

Nevertheless, a key question at the end of this project is what is the appropriate, if any, statistical approach when testing similar effects in the future. Is it a multilevel linear regression? Is it a

CONTACT Angelos-Miltiadis Krypotos amkrypotos@gmail.com

Structural Equation Model (SEM)? Is it perhaps an analysis not even included in Table 1 of Hoogeveen et al. (2022)? Such decisions are critical as researchers, in light of this lack of consensus, may be tempted to select their statistical analysis based on which analysis leads to their preferred result. In the case of *p*-values especially, such analytic flexibility can lead to finding significant results in nearly any dataset, thus inflating the number of false-positives in the literature (e.g., Simmons et al., 2011).

Unfortunately, many analysts' projects are not particularly useful in pitting the different statistical approaches against one another. This is because, as in the present project, the statistical models used often test vague hypotheses featuring theoretically underdetermined constructs (here the relation between "religiosity" and "well-being" and the dependence of the effect on "cultural norms", whatever these are). This two-fold theoretical vagueness engenders two related problems. First, it is unclear how each construct should be operationalized: there are multiple plausible options for each construct in the present dataset. Secondly, different data analytical techniques make different theoretical assumptions: in the absence of clearly articulated assumptions, it is unclear which—multilevel linear regression, SEM, or some other alternative—is most appropriate for testing relations between variables. As such, projects that test vague hypotheses with underspecified constructs will invariably end up in vague, ambiguous, or even ambivalent results. Importantly, this problem is not relevant for many analysts' projects as it is dominant common psychological research.

One solution for drawing statistical inferences in the absence of formal theory is multiverse analysis (Steegen et al., 2016). In this approach, rather than settling on an individual analysis, the analyst defines the full range of reasonable statistical analyses that could be run to determine a result. The researcher then runs all of these analyses simultaneously, and reaches conclusions about the tested hypotheses by considering the results across all of these different analyses (e.g., proportion of *p*-values or Bayes factor below or above a predefined level). Also recently, more sophisticated analyses are suggested for summarizing multiverse effects, such as specification- curves (Simonsohn et al., 2020) where the effects are plotted in a continuous curve so that again researchers can evaluate the robustness of their results. A multiverse approach is currently being proposed in multiple research areas (e.g., fear conditioning; Lonsdorf et al., 2021). Multiverse analyses allow researchers to both test the robustness of their effect to different analytic models, or in the case of a data multiverse, test how different data transformations may also influence the direction of a result when a single model is tested.

Despite the benefits of a multiverse approach, there are three disadvantages. First, multiverse analyses are time consuming. However, this is a lively area of software development and multiple packages are being introduced to mitigate this problem (e.g., Sarma, 2021). Although we agree that whenever running a multiverse approach, maybe running all the different analyses will need more time than running a single analysis, making stronger inferences based on the results worth it.

The second limitation is that a multiverse approach is not a substitute for a formal theory. The need for formal modeling in psychology has been noted decades ago (e.g., Bush & Mosteller, 1951; Meehl, 1978, 1990a, 1990b) and there is a recent rise in awareness of the need of such models in recent literature (e.g., Lewandowsky & Farrell, 2010; Millner et al., 2020). We totally agree with this perspective. The possibility of the multiverse approach should not be taken as an excuse to neglect formal theorizing. Our recommendation of such an approach is relative to its alternative, which is to draw inferences based on single analyses. It may be considered an additional tool in this regard, as other practices also aim to combat selectively, choosing between analyses based on results (e.g., pre-registration, Simmons et al., 2021).

Lastly, it is often difficult to decide which statistical analyses should be included in the multiverse; as always, there could be counter-argument for (not) including a specific statistical approach.

A way to amend this problem is by forming group of expert panels to decide which analyses to include and afterwards the proposed analyses may be evaluated by the community.

Disclosure statement

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

References

Aczel, B., Szaszi, B., Nilsonne, G., Van den Akker, O., Albers, C. J., van Assen, M. A. L. M., … Wagenmakers, E. (2021, April 21). Consensus-based guidance for conducting and reporting multi-analyst studies. https://doi.org/10.31222/osf.io/5ecnh.

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., & Wagenmakers, E. J. (2018). Estimating across-trial variability parameters of the diffusion Decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, *58*(5), 313–323. https://doi.org/10.1037/h0054388

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., & Donkin, C. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069. https://doi.org/10.3758/s13423-017-1417-2

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

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. SAGE.

Lonsdorf, T., Gerlicher, A., Klingelhöfer-Jens, M., & Krypotos, A. M. (2022). Multiverse analyses in fear conditioning research. *Behaviour Research and Therapy*, 104072. doi.org/10.1016/j.brat.2022.104072

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195

Millner, A. J., Robinaugh, D. J., & Nock, M. K. (2020). Advancing the understanding of suicide: The need for formal theory and rigorous descriptive research. *Trends in Cognitive Sciences*, *24*(9), 704–716. https://doi.org/10.1016/j.tics.2020.06.007

Sarma, M. A. (2021). Package 'multiverse'. https://cran.irsn.fr/web/packages/multiverse/multiverse.pdf.

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C., Van Assen, M. A., & Schulte-Mecklenbeck, M. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/j.obhdp.2021.02.003

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, *526*(7572), 189–191. https://doi.org/10.1038/526189a

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration: Why and how. *Journal of Consumer Psychology*, *31*(1), 151–162. https://doi.org/10.1002/jcpy.1208

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637