

Cognitive and cultural models in psychological science: A tutorial on modeling free-list data as a dependent variable in Bayesian regression

Theiss Bendixen^a & Benjamin Grant Purzycki^a

^a *Department of the Study of Religion, Aarhus University, Denmark*

Abstract

Assessing relationships between culture and cognition is central to psychological science. To this end, free-listing is a useful methodological instrument. To facilitate its wider use, we here present the free-list method along with some of its many applications and offer a tutorial on how to prepare and statistically model free-list data as a dependent variable in Bayesian regression using openly available data and code. We further demonstrate the real-world utility of the outlined workflow by modeling within-subject agreement between a free-list task and a corollary item response scale on religious beliefs with a cross-culturally diverse sample. Overall, we fail to find a statistical association between these two instruments, an original empirical finding that calls for further inquiry into identifying the cognitive processes that item response scales and free-list tasks tap into. Throughout, we argue that free-listing is an unambiguous measure of pools of cognitive and cultural information and that the free-list method therefore has broad potential across the social sciences aiming to measure and model individual-level and cross-cultural variation in mental representations.

Keywords: cognitive anthropology, cross-cultural psychology, free-listing, item response scale, zero-one inflated beta regression

1 Introduction

Assessing relationships between culture and cognition is central to psychological science. This is further emphasized by the recent surge in cross-cultural behavioral and psychological research (e.g., Barrett, 2020; Broesch et al., 2020). Developing and applying valid measures of inter-individual and inter-group variation in cognitive and cultural models—including values, attitudes, and beliefs—is therefore critical (D’Andrade, 1987, 1995).

To that end, item response scales are ubiquitous across the social sciences. However, item response scales are not unproblematic¹. For instance, psychometricians have long noted that arbitrary structural features of scales such as the number, order, and difficulty of response options and items systematically influence participants’ response patterns (e.g., Haaf, Merkle, & Rouder, 2020; Krosnick & Presser, 2009; Panter, Tanaka, & Wellens, 1992; Şahin, 2021). In cross-cultural settings, item response scales have also proven difficult to administer in non-numerate field sites (Purzycki et al., 2016a; Purzycki et al., 2022; Quinlan, 2017). Further, while cognitive anthropologists have continually emphasized the importance of obtaining cross-task and construct validity using a diversity of instruments (e.g., D’Andrade, 1995; Dressler, Borges, Balieiro, & Santos, 2005; Furlow, 2003; Romney, Weller, & Batchelder, 1986), item scales are rarely validated using alternative methods of measurement.

One such alternative method is free-listing. In a free-list task, participants are simply invited to list their associations on some topic. The list can—but need not—be capped at some limit. The listed items are then thematically coded according to a specified level of granularity adapted to the particular research goal at hand. Items listed earlier are typically easier to access or recall, and thus constitute a form of *cognitive* salience (Chaves et al., 2019; Romney & d’Andrade, 1964). Further, given a valid topic, items higher up on an individual’s list tend to also be more prevalent across lists drawn from the same sample. This makes the free-list task an unambiguous measure of cultural information (e.g., Bendixen, Apicella, et al., n.d.; Purzycki & Bendixen, 2020; Purzycki, Pisor, et al., 2018; Quinlan, 2017; cf., D’Andrade, 1995, pp. 212–217 for a general discussion of ‘cultural consensus’). These properties make free-listing a promising measure of inter-individual and cross-cultural cognitive models in psychological science.

While free-list data lend themselves to a host of analytic approaches (e.g., Quinlan, 2017; Smith, 1993), free-lists are often mainly used for descriptive or exploratory purposes or for informing downstream item scale construction. Free-listing has seen widespread use in a range of sub-disciplines within cognitive anthropology and psychology (for an overview, see Bernard, 2018, pp. 235–238), including – but not limited to – ethnobiology (e.g., Chaves et al., 2019; Fremout et al., 2021; McNamara & Wertz, 2021; Quinlan, 2005), semantics (e.g., Henley, 1969; Smith & Borgatti, 1997; Smith, Furbee, Maynard, Quick, & Ross, 1995), as well as the study of spiritual and theistic beliefs and practices (e.g., Atran et al., 2002; McNamara, Senanayake, Willard, & Henrich, 2021; Purzycki, 2011, 2013; Purzycki & Holland, 2019; Purzycki & Kulundary, 2018; Singh & Henrich, 2020; Singh, Kaptchuk, & Henrich, 2021; Turpin & Stanford, 2020; Willard, Baimel, Turpin, Jong, & Whitehouse, 2020). But

¹For a principled approach to measuring cultural models using item response theory, see Bunce and McElreath (2017).

recent work has demonstrated the promise of more principled statistical modeling of free-list data, for instance as an independent variable in regression models (e.g., Purzycki, Pisor, et al., 2018; White & Norenzayan, 2021) and in semantic (e.g., Levine, Muthukrishna, Chan, & Satterfield, 2015) and social [e.g., `redhead_mcelreath_ross_2021`] network analysis. Some studies employ free-list data as a dependent variable and, for instance, compare the number of items listed between groups (e.g., Brewer, Garrett, & Rinaldi, 2002) or model the presence of some target item (e.g., Purzycki & Bendixen, 2020; Purzycki, Stagnaro, & Sasaki, 2020; White & Norenzayan, 2021) as a function of some predictor(s) or following experimental manipulation.

When research questions pertain to the psychological salience of some particular concept or item, coded free-list data can be converted into *salience scores* (Quinlan, 2017; Smith, 1993; Smith & Borgatti, 1997). Salience scores range from 0 (i.e., target item not listed) to 1 (i.e., target item listed first) with intermediate values representing the target item's order of listing (i.e., a higher score means that the target item is listed earlier), taking into account the total number of items that a participant lists (see Part Two for formula). Therefore, while modeling presence and absence of a target item might be justifiable on certain theoretical grounds, it is practically akin to dichotomizing a continuous variable and hence discarding information (e.g., Ragland, 1992). At the same time, distributions of data bound by 0 and 1, and where 0 and 1 are valid values, are generally not compatible with common assumptions of popular statistical procedures (Cribari-Neto & Santos, 2019; Douma & Weedon, 2019; Kubinec, 2020; F. Liu & Eugenio, 2018; F. Liu & Kong, 2015; P. Liu, Yuen, Wu, Tian, & Li, 2020; Menezes, Mazucheli, & Bourguignon, 2021; Ospina & Ferrari, 2012; Queiroz & Lemonte, 2021; Santos & Bolfarine, 2015). This poses challenges to analysts who want to model free-list salience scores as a dependent variable and calls for general guidelines on how to go about modeling salience scores and similar zero-one inflated and bounded data. We offer one such set of guidelines here.

This paper has two main parts. In Part One, we outline how to calculate and statistically model salience scores from a free-list task as a dependent variable in Bayesian regression (for introductions to Bayesian data analysis, see Gelman, Carlin, Stern, & Rubin, 2013; A. A. Johnson, Ott, & Dogucu, 2022; Kruschke, 2014; Kruschke & Liddell, 2018; McElreath, 2020). Specifically, we suggest the zero-one inflated beta distribution as likelihood model and illustrate its utility compared to popular alternatives using simulated synthetic data. We also discuss approaches to modeling free-list data without transforming the data into salience scores.

In Part Two, we apply the presented tools in a real-world case study, including steps in how to prepare coded free-list data for modeling. The motivating case study pertains to individual-level data on the moral concerns ascribed to locally relevant deities, a central topic in the cognitive and evolutionary study of religion (for a recent review, see Bendixen, Lightner, & Purzycki, n.d.). Specifically, we model within-subject agreement between a free-list task and a corollary item response scale in a cross-culturally diverse sample and overall, we fail to find a statistical association between these two instruments. This original empirical finding calls for further inquiry into identifying the cognitive processes that item response scales and free-list tasks reflect. We conclude with a call for psychological researchers to consider the free-list task as an alternative or complimentary instrument to item response

scales for measuring cognitive and cultural models of mental representations.

To render the tutorial maximally useful, all code is written in the popular open-source programming language R (R Core Team, 2021, version 4.1.2), and we integrate key code chunks into the main text throughout. Modeling is implemented using the `brms` package (Bürkner, 2017, 2018, 2021), an interface to the probabilistic programming language `Stan` (Carpenter et al., 2017; Stan Development Team, 2021), and free-list data preparation is handled via `AnthroTools` (Jamieson-Lane & Purzycki, 2016; Purzycki & Jamieson-Lane, 2016). Data wrangling and plotting are primarily facilitated by the `tidyverse` (Wickham et al., 2019), `ggplot2` (Wickham, 2016), `tidybayes` (Kay, 2022), and `bayesplot` (Gabry & Mahr, 2021; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019) packages². Data and code to reproduce the full manuscript and appendices, written in R Markdown (Allaire et al., 2021; Xie, Allaire, & Golemund, 2018), are publicly available³.

Part One: Simulating and modeling zero-one inflated and bounded data

1.1 Saliency calculations

A listed item’s *saliency score* depends on both its listed order and the total number of listed items per participant. To calculate a saliency score, an item’s order number, k , is subtracted from 1 plus the total number of items a participant listed, n . This number is then divided by the total number of items listed, $\frac{n+1-k}{n}$. All items listed first thus get an item saliency of 1. Items listed earlier have a higher saliency score (assuming equal list lengths) and are typically easier to access or recall (e.g., Chaves et al., 2019; Quinlan, 2017; Smith, 1993; Smith & Borgatti, 1997). Saliency scores thus reflect a form of *cognitive* saliency. For saliency calculations, we use the `AnthroTools` package (Jamieson-Lane & Purzycki, 2016; Purzycki & Jamieson-Lane, 2016) with the `MAX_SALIENCY` table type function, which takes the maximum saliency of the listed items in cases, where subjects listed the same item more than once. This is particularly useful when participants can list multiple items that would subsequently be coded similarly (e.g., in the context of our case study, ‘theft’, ‘murder’, and ‘deceit’ might all be coded as ‘Morality’; see below and Bendixen, Apicella, et al. (n.d.) for details). These individual-level saliency scores can then be modeled as the dependent variable as a function of some predictor(s). We illustrate a practical implementation in R in Part Two.

1.2 The zero-one inflated beta distribution

Since saliency scores range continuously from 0 to 1, and since 0 (item not listed) and 1 (item listed first) are valid values, we suggest modeling the outcome with a zero-one inflated beta (ZOIB) likelihood distribution with conventional link functions (for formal details, see F. Liu & Eugenio, 2018; F. Liu & Kong, 2015; Ospina & Ferrari, 2012). The ZOIB is a mixture model determined by four parameters: the mean and the precision of the beta distribution;

²See Appendix D for a list of R packages, their dependencies, and version number used for this project, created using `grateful` (Rodríguez-Sánchez, Jackson, & Hutchins, 2022).

³The package `renv` (Ushey, 2022) ensures a reproducible package environment.

a ‘zero-one inflation’ parameter (i.e. the probability that an observation is either 0 or 1); and a ‘conditional one inflation’ parameter (i.e. the probability that, given an observation is 0 or 1, the observation is 1). This specification captures the entire range of possible values while still being constrained between zero and one.

Other common approaches to modeling 0 to 1 bounded outcome data include Gaussian or a regular beta distribution. However, among other limitations, the former depends on the assumption of normality, an assumption that rarely holds in real-world data of this kind and is therefore unlikely to provide a good fit to data in cases where the density of the distribution is not symmetrically centered around a midpoint. Further, if the Gaussian distribution is not truncated, the model can predict values outside the valid range of 0-1. On the other hand, while the beta distribution is more flexible than the normal distribution in that it can capture skewed and non-Gaussian densities, 0 and 1 are not valid values, and hence it is an inappropriate likelihood function for zero and/or 1 inflated data. One proposed solution to this problem is the ‘transformed beta’ distribution, where 0s and 1s are transformed to be, respectively, just above 0 (e.g., 0.001) and below 1 (e.g., 0.999). However, transforming data is arguably unprincipled and fails to capture a key feature of the data-generating process, particularly if the probability of zeros and/or ones are of theoretical interest. In the next section, we informally compare the fit of ZOIB to a Gaussian and transformed beta model using simulated zero-one inflated and bounded data.

An alternative likelihood model for this kind of data is the recently proposed ‘ordered beta distribution’ (Kubinec, 2020). Compared to the ZOIB, which splits the ‘effect’ estimates into several parameters, the ordered beta returns a single ‘effect’ estimate in that it uses the same predictive model for both discrete and continuous outcomes. However, if the probability of zeros and/or ones has a natural interpretation (as is the case for free-list salience scores, where 1 reflects ‘item listed first’ and 0 equals ‘item not listed’) and is of theoretical interest, the ZOIB might still be preferable. In Appendix A, we demonstrate an implementation of the ordered beta and, in the context of the case study of the present paper at least, the ZOIB and the ordered beta yield similar results and posterior predictions.

1.3 Synthetic data and analysis

To illustrate the utility of modeling salience scores (and similar data) using the ZOIB, we simulate and analyze zero-one inflated and bounded data, briefly discuss the interpretation and some simple post-processing of the output from `brms`, and compare the model to alternative likelihood models, namely the Gaussian distribution and the transformed beta. We finally discuss some limitations to the ZOIB and briefly consider some alternative modeling options that do not rely on transforming free-list data into salience scores.

1.3.1 Simulation. We simulate zero-one inflated and bounded data from a ZOIB for two groups, and model the ‘effect’ of group on the distribution of the outcome variable. Imagine, for instance, that these are salience scores from a free-list task administered at two field sites. We are then interested in whether a target item is listed earlier (i.e., is more salient) at one field site compared to the other. The data is stored in `sim_dat`. Assuming

170 default priors (see Part Two for some explication on prior choice), this simple model is
 171 implemented in **brms** as follows⁴.

```
sim_mod <- brm(formula = bf(
  y ~ 1 + group,
  phi ~ 1 + group,
  zoi ~ 1 + group,
  coi ~ 1 + group),
  data = sim_dat,
  family = zero_one_inflated_beta())
```

172 Recall that we are predicting the outcome with four parameters. Here, **y** represents
 173 the mean (**mu**) of the beta distribution. In **brms**, unlike the remaining parameters which
 174 have set names, the mean parameter does not have a default name but will be named
 175 however the dependent variable is named in the data frame. **phi** represents the precision of
 176 the beta distribution, **zoi** is the zero-one inflation, i.e., the probability of zero or one, and
 177 **coi** is the conditional one inflation, i.e., the probability of one given a zero or a one. Here,
 178 we specify our sole predictor, **group**, together with an intercept **1 +**, on all four parameters,
 179 such that **group** can influence all parts of the model, but in principle one could add either
 180 fewer or additional predictor terms to one or more of the parameters. **bf()** facilitates the
 181 specification of several sub-models within the same **formula** call. We then pass **sim_dat**
 182 to the **data** call and set the model **family** to the zero-one inflated beta distribution, which
 183 is native to **brms**. Implicit in the zero-one inflated beta distribution are the link functions,
 184 **mu = logit; phi = log; zoi = logit; coi = logit**, which allows the fitting of linear
 185 models to non-linear relationships. They are important to keep in mind when inspecting
 186 the output. All of this is wrapped within **brm()**, the workhorse of the **brms** package. The
 187 output (not shown directly here) can be summarized with **summary(sim_mod)**.

188 In the summary output, **Population-Level Effects** holds the model estimates
 189 alongside some chain diagnostics, which appear acceptable (a thorough discussion of diag-
 190 nostic metrics of Bayesian models is outside the scope of this tutorial but see e.g., Gelman
 191 et al., 2013; Kruschke, 2014; McElreath, 2020). To get the estimates on their original scales,
 192 however, we need to apply their inverse link functions. Since **mu**, **zoi**, and **coi** are all in
 193 log odds, the logistic function will transform these estimates to their original scales, namely
 194 the mean of the beta distribution and the probability of zero and/or one, respectively,
 195 whereas the **phi** parameters need to be exponentiated due to the applied log link. Ideally,
 196 these transformations should be applied to draws from the posterior distribution and not
 197 on the estimates provided by the **summary()** function, which would amount to summarizing
 198 a summary. Table 1 reports the estimates on their original scales.

199 The **_Intercept** coefficients are the intercepts for each of the sub-models (i.e., the
 200 estimates for group A), while the **_groupB** terms are the differences between group A and B
 201 for each sub-model. A cursory inspection reveals group to have a positive association with
 202 both the mean of the beta distribution **groupB** and the conditional one-inflation **coi_groupB**
 203 in that both estimates and their intervals are solidly above 0.5.

⁴Parts of the code for this section is modified after Vuorre (2019).

Table 1

Parameter estimates of 'sim_mod' after applying inverse link functions to posterior draws.

Parameter	Estimate	Est.Error	Q2.5	Q97.5
b_Intercept	0.26	0.02	0.23	0.30
b_phi_Intercept	4.41	0.53	3.42	5.51
b_zoi_Intercept	0.53	0.03	0.47	0.59
b_coi_Intercept	0.65	0.04	0.57	0.73
b_groupB	0.69	0.02	0.64	0.74
b_phi_groupB	0.73	0.03	0.67	0.79
b_zoi_groupB	0.42	0.04	0.34	0.50
b_coi_groupB	0.67	0.07	0.53	0.79

Note. Point estimates are posterior means.

As is often the case with generalized linear models and in particular in mixture models, model results are easier to inspect when plotted. Figure 1 plots the raw data points (light and dark blue, slightly jittered to prevent over-plotting) and densities for each group together with model predictions of expected values within a 95% interval, taking into account all parameters excluding `phi`. It appears from the distribution of raw data that group B has fewer zeros, a higher mean, and more ones, which is consistent with the model predicting higher expected values on average in group B, supporting our inspection of the posterior estimates.

All in all, given sampling and simulation error with a finite sample, these results recover the parameters with which the data were simulated. Specifically, we simulated the intercepts as `zoi = 0.5`, `coi = 0.65`, `mu = 0.25`, and `phi = 5`, and group B to have a 0.2 positive difference compared to group A on both the mean of the beta distribution and the conditional one inflation, but no association between group and the precision or the zero-inflation.

Finally, we can compare the fit of the ZOIB to simulated data with two alternative likelihood models, the Gaussian distribution and the transformed beta (cf., discussion above). Figure 2 plots the marginal distribution of simulated outcome data (dark blue line) against 50 posterior predictive draws (light blue lines) from the three different likelihood models fitted to the simulated data. The x -axis represents the possible range of outcome values and the y -axis represents the density of each outcome value. Ideally, the predictive draws should show reasonable resemblance with the observed data. The ZOIB clearly exhibits the best fit of the three models, with the Gaussian over-shooting the mean of the simulated data but under-shooting the number of zeros and ones, and the transformed beta in turn under-shooting the mean, over-shooting extreme values and, critically, drops off just before zero and one.

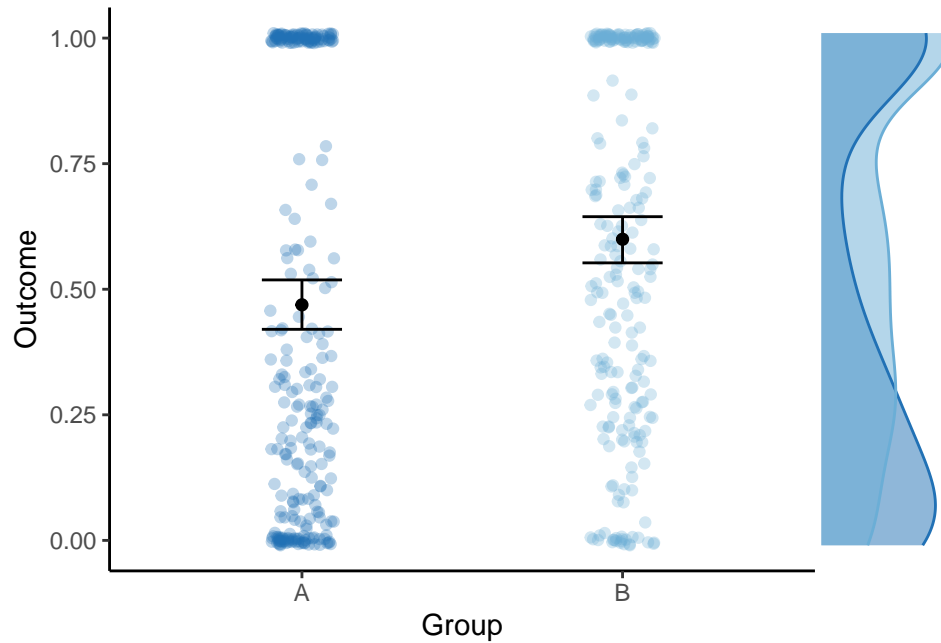


Figure 1. Plot of simulated data and model predictions of expected values. Raw data points (light and dark blue, slightly jittered) and densities from the two simulated groups. Posterior medians (black points) with 95% credible intervals.

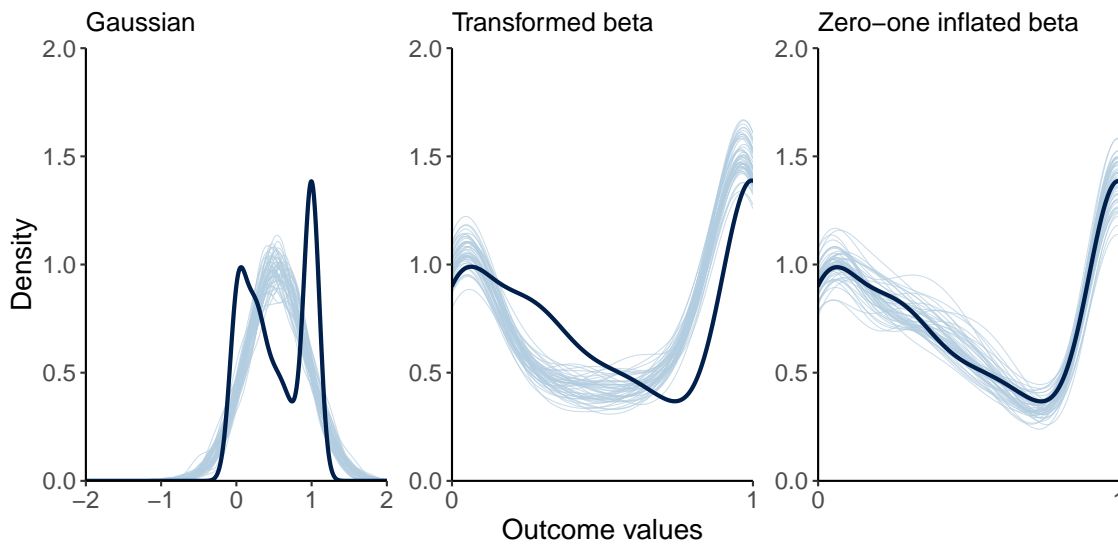


Figure 2. Marginal distribution of simulated outcome data (dark blue lines) against posterior predictive draws (light blue lines) from three different likelihood models fitted to the simulated data.

1.3.2 Discussion. The previous section illustrated the basic implementation and utility of the ZOIB in the context of zero-one inflated and bounded data, such as free-list salience scores. However, by no means have we demonstrated that the ZOIB is superior to all alternatives under all circumstances. For instance, our walk-through was based on a single simulation run with a fixed set of parameters yielding a somewhat unusual distributional shape. It is not surprising that the ZOIB exhibited the best fit since, after all, the data were simulated from a ZOIB distribution! Our goal here is simply to demonstrate that the ZOIB is a viable option to consider in the context of free-list data and similar zero-one inflated and bounded data and to provide readers with some readily-implementable computational tools. With that said, there are limitations to the ZOIB, including the risk of over-parameterization and, thereby, over-fitting (Kubinec, 2020). Similarly, it is unclear whether the ZOIB plausibly reflect any specific cognitive data-generating processes. For instance, in a free-list task, are the cognitive processes that give rise to zeros and/or ones distinct from that giving rise to intermediate values, as modeled with a ZOIB?

In this tutorial, we focus on modeling free-list data in the form of salience scores, as they are a popular transformation of free-list data that simultaneously takes into account both the position of a target item and the total number of items listed per participant while retaining an intuitive interpretation. However, it is of course possible to model free-list data more directly, as for instance modeling the presence/absence of a target item. This is straightforward with a logistic (Bernoulli) regression. Further, using a binomial distribution, researchers can model the number of times a participant lists certain items of interest (i.e., the number of ‘successes’) given the total number of items listed (i.e., the number of ‘trials’), which in turn can be allowed to vary across participants. Alternatively, researchers could model the expected list position of a target item using a zero-inflated negative binomial⁵. The dependent variable would then be the position at which a target item appears for each participant. Similarly, researchers could model the cumulative probabilities of a target item appearing in each of the positions in which the item appears across the sample using zero-inflated ordinal regression. In the latter two cases, a zero-inflation parameter would model the probability of zero (i.e., target item not listed), similarly to the ZOIB, while in the non-zero part of the model a larger regression coefficient for any predictor(s) would indicate that the target item conditionally moves further down participants’ lists (i.e., the target item becomes less ‘salient’). In Appendix B, we illustrate an implementation and some simple post-processing for each of these options. Whether a particular modeling approach is preferable, in general or specifically, will depend on the data material and research question at hand, and we encourage researchers to experiment with these likelihood models—and others that we might have overlooked⁶.

⁵Thanks to Matti Vuorre for this suggestion, as well as for a general comment on keeping the data-generating process in mind when modeling free-list data.

⁶All of the considered models — as well as the model families of which they form a small part — are off-the-shelf options that are increasingly available in contemporary statistical software, and they translate into a very wide range of powerful, practical applications. Eventually, however, analysts could aim to build more ‘bespoke’ models of free-list data, that is, models that are catered to the actual assumed cognitive processes at play. As such models by their very nature sacrifice generalizability for context-specificity, they are necessarily outside the scope of the present tutorial.

PART TWO: Real-world case study

We now turn to an original real-world application, including how to practically prepare coded free-list responses for modeling. The case study stems from the cognitive and evolutionary study of religion and pertains to individual-level data on the moral concerns of locally relevant deities from across seven diverse field sites. Much contemporary research focus on the moral concerns of deities on the grounds that moralizing deities might co-evolve with expanded cooperation (e.g., Botero et al., 2014; D. D. P. Johnson, 2016; Norenzayan et al., 2016; Purzycki et al., 2016b; Purzycki, Henrich, et al., 2018; Schloss & Murray, 2011; Watts et al., 2015). This focus in turn compels researchers to employ valid measures of cross-cultural, individual-level representations of morality and god concepts. As in many other corners of the social sciences, item responses scales are popular instruments to this end (e.g., Atkinson & Bourrat, 2011; Ge, Chen, Wu, & Mace, 2019; Purzycki et al., 2022; Simpson, Piazza, & Rios, 2016).

Here, we conduct a methodological comparison between an item response scale assessing the moral concern of two locally relevant deities and a free-list task on what angers said deities. Overall, we find that the item response scale does not clearly predict corresponding moral concerns in the free-list task. This novel empirical result raises important questions about current methodological approaches in the cognitive and evolutionary study of religion and calls for further inquiry into identifying the cognitive processes that item response scales and free-list tasks tap into. We then discuss candidate methodological explanations for the null result and highlight key implications for the study of moralizing religious traditions. Finally, we argue more generally that the free-list method has broad potential across the social sciences with regards to how researchers measure and model individual-level and cross-cultural variation in mental representations (e.g., beliefs, attitudes, values). For the sake of the tutorial, we again integrate code in the text. The following sections contribute to the tutorial by demonstrating a practical implementation of the ZOIB in a more realistic data setting.

2 Methods

2.1 Data

The present data is publicly available and originates from The Evolution of Religion and Morality Project⁷. The main study consisted of a battery of demographic and religiosity questions as well as behavioral economic games (for further methodological details and presentations of field sites, see Bendixen, Apicella, et al., n.d.; Lang et al., 2019; Purzycki et al., 2016b, 2016a; Purzycki, Henrich, et al., 2018). Specifically for present purposes, participants were asked about the concerns of two different locally relevant deities (a ‘moralistic’ deity and a ‘local’ deity) using both a three-item scale and a free-list task. Table 2 provides an overview of the field sites, their main economies, and the selected deities. For site-specific

⁷Full datasets and protocols are publicly available at: <https://github.com/bgpurzycki/Evolution-of-Religion-and-Morality>.

Table 2
Overview of field sites

Site	Main Economy	Moralistic Deity	Local Deity
Coastal Tanna, Vanuatu	Horticulture/Hunting	Christian God	Garden Spirit (<i>Tupunus</i>)
Inland Tanna, Vanuatu	Horticulture/Hunting	<i>Kalpapen</i>	Garden Spirit (<i>Tupunus</i>)
Lovu, Fiji	Wage Labor	Hindu <i>Bhagwan</i>	—
Mauritius	Wage Labor/Farming	Hindu Shiva	Spirit/Soul/Ghost (<i>Nam</i>)
Pesqueiro, Brazil	Wage Labor	Christian God	Virgin Mary
Tyva Republic	Wage Labor/Herding	Buddha-Burgan	Spirit-Masters (<i>Cher eezi</i>)
Yasawa, Fiji	Fishing/Farming	Christian God	Ancestor Spirits (<i>Kalou-vu</i>)

research reports, see Tanna, Vanuatu: Atkinson (2018); Lovu, Fiji: Willard (2018); Mauritius: Xygalatas et al. (2018); Pesqueiro, Brazil: Cohen, Baimel, and Purzycki (2018); Tyva Republic: Purzycki and Kulundary (2018); Yasawa, Fiji: McNamara and Henrich (2018).

2.2 Item response scale

The item response scale took the form: *How important is it for [moralistic/local deity] to punish: [theft/lying/murder]? Responses were on scales of 0 to 4: (0) Not important at all; (1) A little important; (2) Important; (3) Very important; and (4) The most important. Responses are transformed into an index scale ranging from 0 to 1, by dividing each participant's score on the items by the maximum possible score that each participant could have scored, depending on the number of items that each participant actually completed. For instance, for a participant who answered all three items, the maximum possible score is 12, and for a participant who only completed two items, the maximum possible score is 8. Thus, a participant who completed all three questions and got a total score of 6 would get an index score of $6/12 = 0.5$ ⁸.*

2.3 Free-list data and salience calculations in R

In the free-list task, participants were asked to freely list the kinds of things that angers the [moralistic deity/the local deity]. The free-list data were subsequently thematically coded by two independent coders according to both a top-down, twelve-category coding rubric (our 'general' codes) drawn from Purzycki and McNamara (2016), and a finer-grained, bottom-up rubric subjective to each coder [our 'specific' codes; for further details, see Bendixen,

⁸One issue with this approach is that it assumes that, in cases of missing responses, participants are consistent across the three items—e.g., a participant can get a 1 on the index either by selecting (4) on all three items or on just one or two of the items, if failing to respond to the remaining item(s). Another issue is the lack of uncertainty around the index score. There are other—and arguably more principled—ways of handling missing values, such as full Bayesian imputation (e.g., Erler et al., 2016; McElreath, 2020). We refrain from pursuing this here for a couple of reasons. First, this particular scale has been used in previous research (Lang et al., 2019; Purzycki et al., 2016b) where it failed to predict behavioral outcome in economic games, and so we were interested in associating it with an additional instrument, the free-list task. Second, handling missing values in item scales is outside the scope of this tutorial.

Apicella, et al. (n.d.)). According to the general coding rubric, a response was coded as ‘Morality’, when it satisfied the following description: *generalized behaviors that have a benefit or cost to other people (e.g., hurting, being generous, sharing, etc.)*.

Calculating salience is straightforward using the **AnthroTools** package. Here, we calculate item salience for the general codes across what angers the moralistic deity (BGD). The **CalculateSalience()** command from **AnthroTools** takes several arguments. First, a data frame **data** where the free-list data is stored. Each row is an individual’s free-list response on the *n*th order. Next, an **Order** argument which is the column storing a response’s order number. Then, a **Subj** argument which is the column that contains each individuals’ unique ID number. **CODE** is the coded free-list response. For instance, in this case BGD is the column that contains responses for what angers the moralistic deity. The **GROUPING** argument is optional and is used to group responses depending on group, such as participants from different field sites. It’s not strictly required here; however, if participants’ IDs are not unique across field sites but only within sites, then **GROUPING** needs to be specified.

Finally, **Salience** specifies the name of the new column that will store the calculated salience score, in this case **BGD.S**.

```
BGD.FL <- CalculateSalience(data, Order = "Order", Subj = "CERCID",
                             CODE = "BGD", GROUPING = "Culture",
                             Salience = "BGD.S")
```

Then, a little data wrangling needs to happen. What we want is a data set, where each row is a unique participant with their ID number, cultural group, a salience score for the Morality free-list code and their overall response on the moral interest scale for the given deity. There are many ways to go about this—see source code for one example. We can check the structure **str()** of the resulting data frame, where **id** is participants’ ID numbers, **y** is salience score for free-listed Morality, **culture** is cultural group, and **scale** is overall response on the moral interest scale for the moralistic god’s moral concerns.

```
## 'data.frame':   455 obs. of  4 variables:
## $ id      : chr  "CERC1" "CERC10" "CERC11" "CERC113" ...
## $ y       : num   1 1 1 0 1 0 0.5 0 1 0 ...
## $ culture : chr   "Coastal Tanna" "Coastal Tanna" "Coastal Tanna" "Inland Tanna" ...
## $ scale  : num   0.167 0.333 0.417 0.25 0.333 ...
```

3 Modeling salience scores

Recall that we want to assess whether the ‘moral interest scale’ of a focal deity corresponds to salience of moral items in the free-list task. Specifically, if a participant responded on the scale that a focal deity (moralistic or local deity) was morally punitive, both overall on the scale and on sub-items (theft, lying, murder), would corresponding responses—that is, ‘Morality’ as a general code and, as specific codes, theft, lying and murder—also be higher up on a participant’s free-list regarding what angers the focal deity? We construct a series of

models assessing these relationships: 1) for both the moralistic and local deity, 2) between the index score on the scale and free-listed ‘Morality’ (general code), and 3) between each item of the scale (theft, lying, murder) and the salience of each of these focal items in the free-lists. Since we have no reason to suspect systematic missingness patterns, we perform complete-case analysis. Below, we detail our analysis, first in formal notation and then in R code.

3.1 Overview

We model two deities (moralistic and local) and for each of these we model, first, the relationship between free-listed, general-coded ‘Morality’ and the moral interest scale (Model 1 and Model 2) and then the relationship between free-listed, specifically-coded moral items (theft, lying, murder) and the corresponding single items of the moral interest scale (Models 3–8). This yields eight models. As we cluster by field site, these are multilevel models, and for each main model, we also fit a corresponding ‘null’ model that excludes scale response as a predictor of the free-list outcome but retains varying effects of field site, and then employ approximate leave-one-out cross-validation (Vehtari et al., 2020; Vehtari, Gelman, & Gabry, 2017) to assess the relative out-of-sample predictive performance of models with and without scale as a predictor. Throughout model-building, checking and inference, we informally follow a Bayesian workflow (Gelman et al., 2020). In the following sections, we spell out the general formal form of the eight main models, starting with Models 1-2, as well as their implementation in R.

3.2 Formal form of Models 1-2: Predicting general Morality in the free-list task

All our linear models include: global intercepts $\beta^{\text{Intercept}}$, site-specific (indexed by j) intercepts $\beta_{[j]}^{\text{Intercept}}$, global effect for scale score $\beta^{\text{Scale}} S_i$, and varying slopes across sites for scale score $\beta_{[j]}^{\text{Scale}} S_i$. The models thus take the general formal form:

$$\begin{aligned} \text{Free-listed Morality}_i &\sim \text{ZOIBeta}(\mu_i, \phi_i, \alpha_i, \gamma_i) \\ \text{logit}(\mu_i) &= \beta_{\mu}^{\text{Intercept}} + \beta_{\mu,[j]}^{\text{Intercept}} + \beta_{\mu}^{\text{Scale}} S_i + \beta_{\mu,[j]}^{\text{Scale}} S_i \\ \text{log}(\phi_i) &= \beta_{\phi}^{\text{Intercept}} + \beta_{\phi,[j]}^{\text{Intercept}} + \beta_{\phi}^{\text{Scale}} S_i + \beta_{\phi,[j]}^{\text{Scale}} S_i \\ \text{logit}(\alpha_i) &= \beta_{\alpha}^{\text{Intercept}} + \beta_{\alpha,[j]}^{\text{Intercept}} + \beta_{\alpha}^{\text{Scale}} S_i + \beta_{\alpha,[j]}^{\text{Scale}} S_i \\ \text{logit}(\gamma_i) &= \beta_{\gamma}^{\text{Intercept}} + \beta_{\gamma,[j]}^{\text{Intercept}} + \beta_{\gamma}^{\text{Scale}} S_i + \beta_{\gamma,[j]}^{\text{Scale}} S_i \end{aligned}$$

where μ and ϕ represent, respectively, the mean and the precision of the beta distribution, α represents ‘zero-one inflation’ (the probability that an observation is either 0 or 1) and γ represents ‘conditional one inflation’ (the probability that, given an observation is 0 or 1, the observation is 1).

We define our priors, which we checked through iterative cycles of prior predictive simulation (see below), as follows:

$$\begin{aligned}
\beta_{\mu}^{\text{Intercept}} &\sim \text{Normal}(0, 1) \\
\beta_{\mu}^{\text{Scale}} &\sim \text{Normal}(0, 2.5) \\
\beta_{\phi}^{\text{Intercept}} &\sim \text{Normal}(2, 0.5) \\
\beta_{\phi}^{\text{Scale}} &\sim \text{Normal}(-1, 1) \\
\beta_{\alpha}^{\text{Intercept}} &\sim \text{Normal}(0, 1) \\
\beta_{\alpha}^{\text{Scale}} &\sim \text{Normal}(0, 2.5) \\
\beta_{\gamma}^{\text{Intercept}} &\sim \text{Normal}(0, 1) \\
\beta_{\gamma}^{\text{Scale}} &\sim \text{Normal}(0, 2.5)
\end{aligned}$$

These priors are weakly regularizing as they constrain the estimates to be mostly within realistic ranges and are mildly skeptical toward strong statistical relationships. It is important to stress, however, that priors should be interpreted in the context of the data and model of which they are part (Gelman, Simpson, & Betancourt, 2017), and that these priors might therefore not generalize to other modeling scenarios. Developing principled and practical approaches to eliciting appropriate prior distributions is an active research area, and further discussion is outside the scope of this tutorial (for a technical review, see Mikkola et al., 2021; for a practical introduction, see McElreath, 2020).

Except for ϕ , the prior coefficients are on the log odds scale, meaning that to get the coefficients on their original scale, we first turn them into odds by exponentiating them (cf., Part One). The original scale is then odds/(1 + odds). So, for instance, an intercept prior with mean 0 on the log odds scale means that the coefficient is centered on $\exp(0)/(\exp(0) + 1) = 1/(1 + 1) = 0.5$, which is the midpoint of our outcome variable (i.e., saliences scores ranging from 0 to 1). Likewise, since two standard deviations roughly cover 95% probability of a normal distribution, assigning a prior of $\text{Normal}(0, 1)$ sets the bulk of the prior probability mass to be within the range of $\exp(\pm 2)/(\exp(\pm 2) + 1) \approx \pm 7.38/(\pm 7.38 + 1) \approx [0.12, 0.88]$. In this particular parameterization, the β^{Scale} coefficients represent the estimated change in outcome (i.e., salience score) going from zero on the predictor (i.e., the lowest response on the index) to 1 (i.e., the maximum response on the index). A $\text{Normal}(0, 2.5)$ prior is therefore sensible in that it puts some probability mass on the entire range between ≈ -1 and ≈ 1 while being centered on 0 (i.e., no change)⁹.

As we cluster on field site, each linear model $(\mu, \phi, \alpha, \gamma)$ has its own multi-level structure, which is formally defined as follows. The varying effects across sites are bound in a variance-covariance matrix with multivariate priors of a Gaussian variety (with means of 0 as these are already included in the linear models above):

$$\begin{bmatrix} \beta_{\mu, \phi, \alpha, \gamma, [j]}^{\text{Intercept}} \\ \beta_{\mu, \phi, \alpha, \gamma, [j]}^{\text{Scale}} \end{bmatrix} \sim \text{Multivariate Normal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{SRS} \right)$$

⁹Note that when we model these priors in a joint distribution, the prior for the intercepts in the μ , α , and γ sub-models widens so as to allow for the steep slopes implied by the prior on β^{Scale} .

Here, \mathbf{S} is a diagonal matrix of intercept and predictors' standard deviations, σ_p ,

$$\mathbf{S} = \begin{bmatrix} \sigma_{\beta^{\text{Intercept}}_{\mu, \phi, \alpha, \gamma}} & 0 \\ 0 & \sigma_{\beta^{\text{Scale}}_{\mu, \phi, \alpha, \gamma}} \end{bmatrix}$$

$$\mathbf{S} \sim \text{LKJcorr}(4)$$

$$\sigma_p \sim \text{Exponential}(1)$$

while \mathbf{R} is their correlation matrix with a prior distribution from the Lewandowski family (Lewandowski, Kurowicka, & Joe, 2009).

3.3 Implementation in R

All of this is implemented in R with `brms` as follows, illustrated with Model 1 `m1`, which predicts the salience of 'Morality' as a general code using the three-item moral interest scale: First, we specify our model `formula`. In the present study, the model formula is relatively simple, in that we are predicting salience scores (`y`) according to an intercept 1 and a single predictor, the moral interest scale (`scale`), which is clustered according to field site (`culture`), yielding the basic syntax `y ~ 1 + scale + (1 + scale | culture)`. This specification allows each field site to have varying intercepts and slopes. We specify this structure on all four sub-models.

Next, since this is a Bayesian model, we want to explicitly set the prior distributions for each of the model parameters and hyper-parameters (see above). `brms` includes a convenience call, `get_prior()`, that returns a complete list of model parameters and their names given a `formula`, a data frame `data`, and a likelihood function `family`.

Then, we manually specify our priors, using the `set_prior()` command in `brms` and store it in a new object, `priors_main`. Note that this step is not strictly necessary to make the model run. In the absence of user-specified prior distributions, `brms` implements default 'flat' priors that will likely work well in many applied cases and yield results that approximate maximum likelihood estimation.

```
priors_main <- set_prior("normal(0, 2.5)", class = "b") +
  set_prior("normal(0, 1)", class = "Intercept") +
  set_prior("exponential(1)", class = "sd") +
  set_prior("normal(0, 2.5)", class = "b", dpar = "coi") +
  set_prior("normal(0, 1)", class = "Intercept", dpar = "coi") +
  set_prior("exponential(1)", class = "sd", dpar = "coi") +
  set_prior("normal(-1, 1)", class = "b", dpar = "phi") +
  set_prior("normal(2, 0.5)", class = "Intercept", dpar = "phi") +
  set_prior("exponential(1)", class = "sd", dpar = "phi") +
  set_prior("normal(0, 2.5)", class = "b", dpar = "zoi") +
  set_prior("normal(0, 1)", class = "Intercept", dpar = "zoi") +
  set_prior("exponential(1)", class = "sd", dpar = "zoi") +
  set_prior("lkj_corr_cholesky(4)", class = "cor")
```

Next, we verify that these priors are sensible. We aim for so-called weakly regularizing priors, that is, priors that put most of their probability mass within realistic ranges. To do so, we perform prior predictive checks, whereby we sample from the model and prior without fitting the model to the data. Note that we now pass our formula, data, and likelihood to `brm()`, and that we also add our priors via the `prior` call. We set `sample_prior = "only"` so that we only sample from the prior and not the data, and a `seed` for reproducibility.

```
m1priorcheck <- brm(formula = bf(
  y ~ 1 + scale + (1 + scale | culture),
  phi ~ 1 + scale + (1 + scale | culture),
  zoi ~ 1 + scale + (1 + scale | culture),
  coi ~ 1 + scale + (1 + scale | culture)),
  prior = priors_main,
  data = bgd_gen_data,
  family = zero_one_inflated_beta(),
  sample_prior = "only", seed = 2021)
```

There are many options for assessing the implications of the priors [see e.g., Gabry et al. (2019); Kruschke (2014); McElreath (2020)]¹⁰ and oftentimes, prior predictive checks need to be catered to the specific data and model(s) at hand (Gelman et al., 2020). Here, we show just two different readily-available visualization options. First, we can sample from the prior predictive distribution and compare those values to the empirical distribution of observed values, using the convenience function `pp_check()` in the `bayesplot` package (Figure 3, top). We simply pass our prior model fit `m1priorcheck` to `pp_check()` and set the number of samples (say, 30) that we want from the prior predictive distribution `ndraws = 30`, as well as a `seed` for reproducibility. This is the same procedure as we used to generate Figure 2.

The dark blue line in the upper panel of Figure 3 is the distribution of observed values, whereas the light blue lines are draws from the prior predictive distribution. As can be seen, the prior draws mostly fall within realistic estimates (around the observed values) and only a few draws are in extreme ranges (i.e., large numbers of zeros and/or ones). Recall that this is so even before the model has “seen” the data (the `sample_prior = "only"` call makes the model sample only from the priors). Note, however, that we include the observed distribution only for illustration here; ideally, the priors should be sensible *a priori*, that is before either the model *or* the analyst has inspected—or even collected—the data.

While this kernel density plot is a quick way to visualize a model’s prior expectation of the marginal outcome distribution, it gives us nothing in the way of assessing the priors for the influence of the predictor terms; that is, what the model expects the relationship to be between the independent and the dependent variable(s). For that, we can instead plot draws from the prior predictive distribution of the outcome as a function of the predictor (Figure 3, bottom). Each line is a regression line of expected values. As can be seen, some lines are flat, some slope downwards and yet others upwards, suggesting that the model

¹⁰See also Kurz’ (Kurz, 2021a, 2021b) on-going code translations of Kruschke (2014) and McElreath (2020).

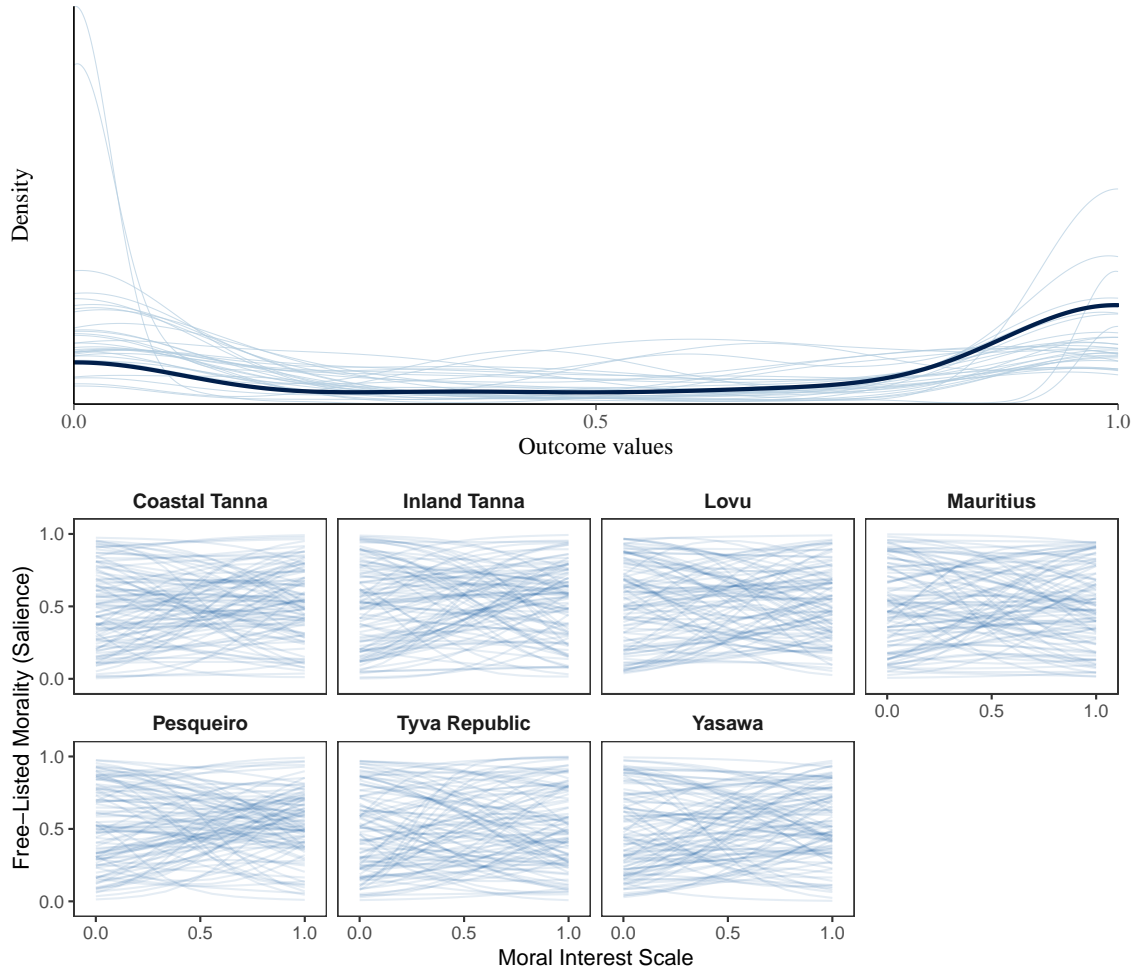


Figure 3. Prior predictive checks. Top: Prior predictions of the marginal distribution of the outcome variable (kernel density). The dark blue line is the distribution of observed values. The light blue lines are draws from the prior predictive distribution. The y -axis represents density, while the x -axis denotes outcome scores. Bottom: Prior predictions for each field site from Model 1 on the relationship between the moral interest scale (x -axis) and saliency of the free-listed general Morality code (y -axis). Lines are draws of expected values from the prior predictive distribution.

allows a wide range of statistical relationships. Finally, we are ready to fit the full model and then perform *posterior* predictive checks, to assess how much the model learned from the data (see Figure 4), compared to these *prior* predictive checks. In the following code block, the `formula`, `data`, `prior`, and `family` calls are similar to above, and we then add some explicit settings to the sampler, which in turn facilitate reasonable chain convergence and model comparison.

```
m1 <- brm(formula = bf(
  y ~ 1 + scale + (1 + scale | culture),
  phi ~ 1 + scale + (1 + scale | culture),
  zoi ~ 1 + scale + (1 + scale | culture),
  coi ~ 1 + scale + (1 + scale | culture)),
  data = bgd_gen_data,
  prior = priors_main,
  family = zero_one_inflated_beta(),

  # some chain controls
  cores = 4, chains = 4,
  iter = 6000, control = list(adapt_delta = 0.99),
  seed = 2021,

  # save_pars facilitates model comparison
  save_pars = save_pars(all = TRUE))
```

Note that the posterior predictive draws in the kernel density plot are much closer to the distribution of observed values, indicating that the model indeed has learned from the data. With `summary(m1)` we can get an overview of the main parameter estimates and chain diagnostics of the model.

3.4 Formal form and implementation of Models 3-8: Predicting specific moral items in the free-list task

The key difference between the general-coded ‘Morality’ models outlined in the previous section (Models 1-2) and the specific moral item models (Models 3-8) is that in the latter set we model the predictor, the single moral scale item, with a monotonic function, since it is an ordered category of response options (Bürkner & Charpentier, 2020; Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018). This allows each response option (from 0 to 4) to have its own conditional effect (though with the same directionality, i.e. monotonicity) on the outcome, which is now salience of a single free-listed item. These monotonic predictors are modeled with a weakly informative prior in the form of a Dirichlet distribution, $\text{Dir}(2, 2, 2, 2)$. This prior encodes the expectation that any of the response options could be more or less likely than the others (McElreath, 2020, pp. 391–396).

The remaining priors are otherwise identical to Models 1-2 with the following exception. Compared to Models 1-2, where the β^{Scale} coefficients for the μ , α , and γ sub-models

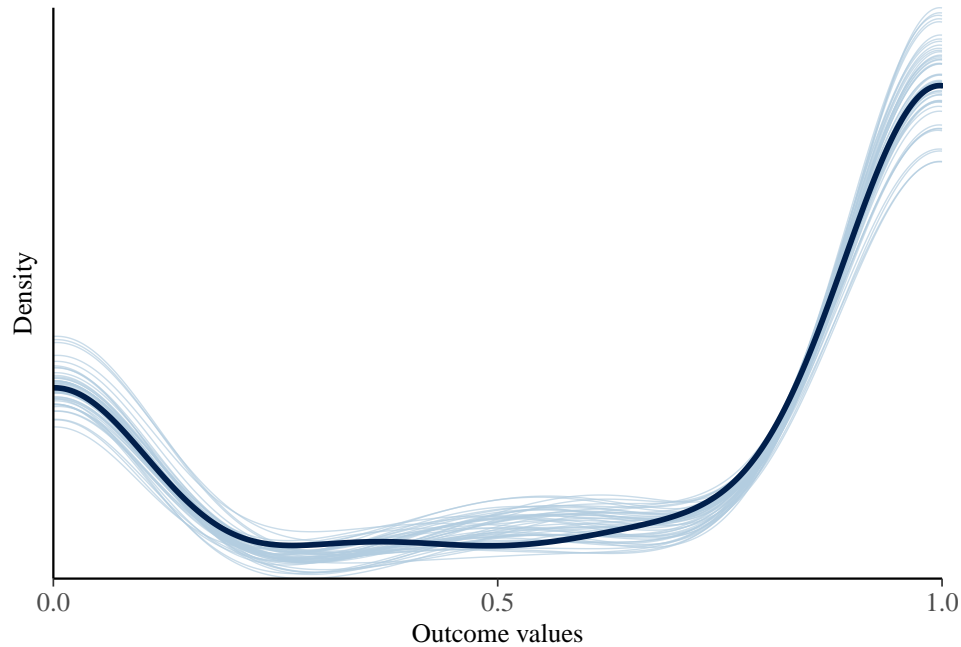


Figure 4. Posterior predictive check for `m1`. The dark blue line is the distribution of observed values. The light blue lines are draws from the posterior predictive distribution. The y -axis represents density, while the x -axis denotes outcome scores.

are interpreted as the estimated change in outcome when the predictor is set at one (i.e., the maximal response on the index) compared to zero (i.e., the lowest response on the index), in Models 3-8 the β coefficients instead represent the estimated average difference between two adjacent categorical response options (cf., Bürkner & Charpentier, 2020). Therefore, we set a prior on the β^{Scale} coefficients for the μ , α , and γ sub-models in Models 3-8 to $\text{Normal}(0, 0.625)$, where the standard deviation of 2.5 from the Models 1-2 β priors the μ , α , and γ sub-models is divided with 4, the number of response options minus 1.

To implement these changes in `brms`, we need only a few modifications to the R code above. To specify monotonicity, we simply wrap the predictor, `scale`, in `mo()`, and `brms` takes care of the rest, and we set a Dirichlet prior on the monotonic predictors similarly to above and add it to the prior call (for further details, see Bürkner & Charpentier, 2020): `set_prior("dirichlet(2,2,2,2)", class="simo", dpar=c("", "coi", "phi", "zoi"), coef="moscale1")`.

4 Results and Discussion

4.1 Results

Here we report key results. First, we assess the posterior predictions of Model 1 and 2 (see Appendix C for similar plots of Models 3-8), and then compare model comparison metrics between all models and their respective ‘null’ models.

Figure 5 plots, for each field site and each deity (top: moralistic deities, bottom: local deities), draws of expected values from the posterior predictive distribution¹¹, as well as raw data points with slight jitter. For the moralistic deities, in most sites the draw lines are mostly flat, suggesting that the relationship between the moral interest scale (x -axis) and salience of the free-listed general Morality code (y -axis) is weak or non-existent. At all sites for the local deities, the relationship seems to trend slightly upwards, although the estimates are quite uncertain, as visualized by the widely-spaced draw lines at several sites. Note that we get posterior predictions for Lovu for the local deities even though we do not have data for a local deity at that particular field site. Rather, the model predicts expected values for a hypothetical unobserved field site, which we label Lovu.

To put these results into perspective, consider the raw data points. Data points in the top-left corner of a panel are individuals who, when asked on the scale how important it is for the target deity to punish theft, lying, and murder, respond ‘not at all important’, but free-list a moral item as the top thing that angers the target deity. And vice versa for the bottom-right corner: here, participants respond that theft, lying, and murder are all the most important thing for the target deity to punish¹² but does not free-list a moral item among that deity’s concerns. That a noticeable number of participants gravitated toward such response patterns is an illustration of the model-driven inferences, namely that these data and model predictions suggest no relationship—and at best a very weak and uncertain one—between the scale and the free-list task. Results are qualitatively similar when we assess the posterior predictions of the specific codes and the corresponding moral scale items (see Appendix C).

This interpretation is further bolstered when we compare each main model against its respective ‘null’ model. Table 3 reports raw (i.e., without having applied their inverse link functions, unlike in the simulated case above) β coefficients (posterior means) with 95% credible intervals for each main model’s focal parameters. If the scale and the free-list task is positively associated, we would expect either of these parameters to have estimates where the bulk of the interval is above zero (with the caveat that the one inflation parameter $\beta_{\gamma}^{\text{Scale}}$ is conditional on the zero-one inflation parameter $\beta_{\alpha}^{\text{Scale}}$). Table 3 also reports model comparison metrics obtained with approximate leave-one-out cross-validation (LOO-CV). In a multilevel context, LOO-CV is an indicator of a model’s accuracy in predicting a new observation in one of the observed groups¹³. LOO-CV is a preferable model comparison metric as it penalizes model complexity. This is desirable for present purposes as we aim to identify the best performing and most parsimonious model within each of the eight model pairs. Akaike weights are calculated on the basis of the leave-one-out information criterion

¹¹More formally, these are draws from the posterior distribution given by (cf., Ospina & Ferrari, 2012): $E[y] = \Pr(\text{zero or one}) \times \Pr(\text{conditional one}) + \text{mean of the beta distribution} \times \Pr(\text{neither zero nor one}) = \alpha \times \gamma + \mu \times (1 - \alpha)$. As can be seen, this calculation ignores the precision parameter of the beta distribution ϕ .

¹²For present purposes, we ignore the logical incoherence of such a response pattern (i.e., responding that two or more distinct behaviors are *the most* important thing to punish).

¹³An alternative approach is leave-one-group out cross-validation (LOGO-CV) which assesses a model’s accuracy in predicting a new observation in a new (unobserved) group. This procedure is more computationally intensive, because it involves refitting the model k times, where k equals the number of observed groups. See Appendix C for an implementation. Results are qualitatively similar to LOO-CV in the present study.

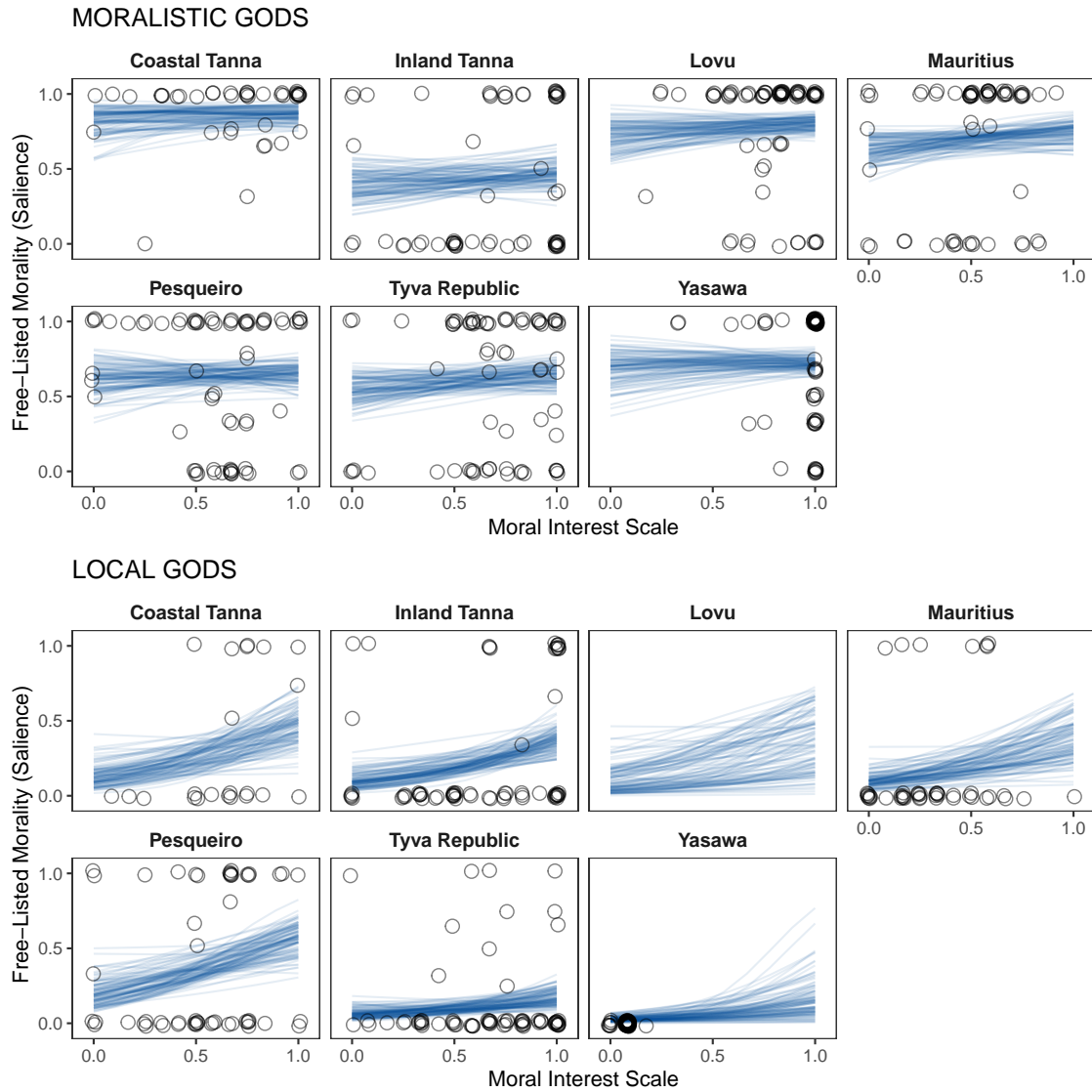


Figure 5. Posterior predictions for each field site and each deity from Model 1 and 2 on the relationship between the moral interest scale (x -axis) and salience of the free-listed general Morality code (y -axis). Lines are draws of expected values from the posterior distribution. Raw data points are slightly jittered.

Table 3

Focal parameter estimates for all models and model comparison metrics.

	$\beta_{\mu}^{\text{Scale}}$	$\beta_{\alpha}^{\text{Scale}}$	$\beta_{\gamma}^{\text{Scale}}$	Akaike Weights	ELPD difference [SE]	N
m1	-0.35 [-0.98, 0.29]	-0.20 [-1.30, 0.90]	0.54 [-0.52, 1.60]	0.10	-1.86 [2.11]	455
m1_null	—	—	—	0.90	*	
m2	1.07 [-0.18, 2.41]	-0.95 [-2.77, 0.88]	1.80 [0.38, 3.28]	0.70	*	326
m2_null	—	—	—	0.30	-0.95 [3.74]	
m3	0.03 [-0.11, 0.17]	-0.02 [-0.25, 0.23]	0.09 [-0.26, 0.42]	0.01	-4.19 [1.91]	451
m3_null	—	—	—	0.99	*	
m4	-0.08 [-0.34, 0.19]	-0.03 [-0.35, 0.28]	-0.42 [-0.95, 0.12]	0.01	-4.25 [1.99]	451
m4_null	—	—	—	0.99	*	
m5	-0.08 [-0.17, 0.13]	0.04 [-0.04, 0.10]	0.09 [-0.22, 0.39]	1.00	*	451
m5_null	—	—	—	0.00	-38.40 [19.44]	
m6	-0.18 [-0.73, 0.35]	0.08 [-0.46, 0.65]	-0.05 [-0.83, 0.60]	0.08	-2.48 [2.08]	326
m6_null	—	—	—	0.92	*	
m7	-0.10 [-0.60, 0.49]	-0.08 [-0.65, 0.53]	-0.54 [-1.33, 0.23]	0.15	-1.82 [1.38]	326
m7_null	—	—	—	0.85	*	
m8	-0.06 [-0.56, 0.48]	0.14 [-0.47, 0.84]	-0.64 [-1.65, 0.41]	0.02	-4.17 [2.41]	326
m8_null	—	—	—	0.98	*	

Note. Point estimates of the raw β coefficients are posterior means with 95% credible intervals in brackets. The best performing model of the eight pairs is denoted by an asterisk. N is the total number of observations used in each model (identical across model pairs).

and a higher value means more weight, i.e. the model that the algorithm more heavily favors. Expected log posterior density (ELPD) is a measure of overall model fit and out-of-sample predictive accuracy. Here, we report the ELPD difference and the standard error (SE) of the difference between the best performing model, denoted by an asterisk, and the remaining model in the pair.

While ELPD differences are generally small compared to their standard errors, in most cases we find that LOO-CV either favors the null models or favors none of the models in particular, indicating that including scale response as a predictor of moral salience in the free-list task does not generally improve out-of-sample predictive performance when penalizing increased model complexity. The only clear exception to this pattern is the Model 5 pair, where the full model (m5) is clearly favored over the null model. Note, however, for this particular model, we had chain convergence issues, potentially rendering estimates unreliable. Further, in an alternative specification of the Model 5-pair using the ordered beta distribution as the outcome likelihood the interval of the β coefficient straddles zero, although here too LOO-CV favors the main model over the null. However, upon inspecting field-site specific estimates this appears primarily driven by the Lovu data (see Appendix C). Overall, then, we infer from these lines of evidence that there is no clear and consistent association between the moral interest scale and the free-list task across gods and field sites.

4.2 Discussion

We now turn to a discussion of, first, possible methodological accounts of the null result, and then consider some of its key implications.

First, many participants did not free-list the target moral items—note the zeros along the x -axis in Figure 5. One potential issue might be that these zeros drive the non-association between the two instruments. However, we claim that excluding participants who did not free-list moral items in this case is misleading. These participants *could* have listed moral items, but they simply did not. Therefore, there is information in the zeros, which may in turn help guide interpretation. On the other hand, since the free-list task was capped at five items (a few participants did list more than that), it stands to reason that participants potentially could have listed the target moral items eventually but that the design did not allow for it. Of course, items listed later would have a lower salience score and so this argument does not account for the—at best—very weak and uncertain relationship between the two instruments. To address this, we re-ran the main models (m1 – m8) with reduced data sets which only included non-zero free-list responses. Again, we found no reliable association between the two instruments across deities and field sites (see Appendix C).

Other methodological limitations should also be considered. It could be, for instance, that the non-existing relationship between the two instruments is simply an artifact of the study protocol. Participants completed the scale before completing the free-list task. Hence, it is possible that participants were simply reluctant to free-list the target moral items (theft, murder, lying) because they had already completed scale items pertaining to these. We are skeptical of this explanation for several reasons. First, this does not explain the non-relationships for moral items in general (Models 1-2), as the Morality code includes moral behaviors in general, not only the three target moral items that were the focus in the item response scale. Second, the protocol was quite long (around 90 minutes), and the free-list task was not preceded directly by the moral items. Therefore, we further deem it unlikely that the scale items had a noticeable effect on the free-list task. Third, this ordering effect cannot account for the disparate responses that participants gave across field sites and deities. Note that there are many more non-zero free-list responses in the panels of the moralistic deities compared to the local deities. In other words, if participants were reluctant to free-list moral items because they had already completed the moral items, why were they then more likely to free-list moral items for the moralistic gods than the local gods? Fourth, the ordering effects explanation contradicts another possible explanation for the results, namely that some kind of priming influenced the free-list task. Just as the order effects could hamper the tendency to report the target items, in principle it is also possible that the scale items could have increased the tendency to report the target items, perhaps through a priming mechanism. We are skeptical of the priming explanation for the same reasons as the reverse. In sum, ordering effects are not plausible explanations in our view.

Lastly, one could object that the instructions for the two tasks are not sufficiently similar to warrant an empirical comparison; the item scale asks about behaviors that a given deity might *punish* whereas the free-list task probes the kinds of things that *angers* a given deity. We concede this objection. However, we do not think that it explains away

the reported results. Ethnographically, there is often considerable overlap between what a deity is angered by and what said deity punishes, at least for these data (Bendixen, Apicella, et al., n.d.) but also in the ethnographic record more generally (e.g., Bendixen & Purzycki, n.d.). And so, while we would perhaps not expect a perfect association between the two tasks, we would at least expect a detectable positive association. Of course, another possibility is that the two tasks measure distinct cognitive processes. This is an intriguing area of future research; however, we refrain from discussing this possibility further here, since our data do not allow us to make inferences one way or the other.

The reported null results speak directly to current debates in the cognitive and evolutionary sciences of religion. A rapidly growing body of work assesses the relationship between moralizing religions and deities and social behavior (Bendixen, Lightner, et al., n.d.) and item response scales are very common instruments as measures of moral salience of deities and religions. However, our finding that an item response scale on a target deity's moral concerns does not predict moral salience of that deity in a free-list task raises questions about current methodological approaches and is consistent with another recent cross-cultural study that failed to find clear evidence of an association between item scales and free-list tasks on moral and political attitudes (White & Norenzayan, 2021).

To complicate matters further, in previous cross-cultural analyses on a data set of which the present one is a subset, the moral interest scale failed to predict individual-level behavior in two different economic games involving more than 2000 participants from 15 diverse field sites (Lang et al., 2019). This finding casts doubt on either a) the hypothesis that moral salience of a deity *per se* is associated with cooperative tendencies on the individual level, b) the prospect of using item response scales to capture individual-level variation relevant to this domain, or c) both. An additional downside to item response scales identified by this cross-cultural research is that Likert scales are not intuitive for some peoples relatively isolated from Western influence (cf., Apicella, 2018; Purzycki et al., 2016a; Purzycki et al., 2022), a limitation that does not apply to oral free-listing (Quinlan, 2017).

In summary, in contrast to pre-fabricated item response scales, free-listing is an unambiguous measure of pools of cognitive and cultural information; participants drawn from the same sample independently tend to exhibit convergence with regard to their mental models, given a valid topic (see e.g., Bendixen, Apicella, et al., n.d.; Purzycki & Bendixen, 2020; Purzycki, Pisor, et al., 2018; Willard et al., 2020). Free-listing is also arguably a more natural analog to how people in their day-to-day lives socially express, transmit, and exchange mental representations. For these reasons, we encourage researchers interested in measuring cognitive and cultural models to include alongside an item response scale a corresponding free-list task so as to ensure that their scales have explanatory power above and beyond a measure of task-specific item response behavior.

5 General Discussion and Conclusion

Here, we have 1) offered a tutorial on how to prepare and statistically model salience scores from a free-list task in Bayesian regression with various predictor terms (i.e., a binary predictor in Part One and a continuous and a categorical predictor in Part Two, respectively)

using readily-available computational tools, and 2) demonstrated the real-world applicability of this workflow by empirically assessing within-subject agreement between a free-list task and an item response scale on religious beliefs with a cross-culturally diverse sample. We failed to find evidence of a clear and consistent relationship between these two instruments, and discussed some possibilities as to what might explain this finding, at least in the particular domain of religious beliefs. At the very least, the presently reported results call for further research into the cognitive processes underlying item response scales and free-list tasks and pose an exciting challenge for field researchers, psychologists, and cognitive scientists to solve in future collaborative efforts. Another potential avenue for future research is to develop a principled and broadly implementable workflow for handling missing values in free-list tasks, accounting for uncertainty in an item's list position, as well as relating individual-level probability distributions of responses to a target population of interest (cf., Deffner, Rohrer, & McElreath, 2021).

Further, in terms of how to measure and model individual-level and cross-cultural variation in mental representations (e.g., beliefs, attitudes, values), we want to stress that the present work has ramifications that reach beyond the scholarly field in which the presented case study was conceived. For instance, there is arguably untapped potential in employing free-list tasks in measuring not only religious beliefs, but also beliefs and attitudes on science (e.g., Bendixen, 2020), morality (e.g., Purzycki, Pisor, et al., 2018; White & Norenzayan, 2021), health (e.g., Quinlan, 2017), as well as political, ideological, and economic issues (e.g., Bendixen, 2019; White & Norenzayan, 2021), disciplinary areas in which item response scales are extremely common but oftentimes not externally or cross-task verified. Here too, the free-list task could serve as a convenient measure of cross-task robustness or as a methodological tool to explore whether these two tasks indeed tap into distinct cognitive processes. As such, the original analysis reported herein is synergistic with the scope of the outlined tutorial; the reported null results emphasize a heightened need for attempts at cross-task verification of item response scales across disciplines. This in turn highlights the inevitability of engaging in detailed and rigorous ethnographic¹⁴ data collection and, thereby, motivates the dissemination of principled and implementable guidelines on how to model such data, including free-list outcomes, as alternative or complimentary instruments to item response scales.

More generally, zero-one bounded and inflated data are commonplace in many applied contexts. For instance, social scientists employ visual slider scales, 'feeling thermometers', and other (semi-)continuous indices whereby participants rate their sentiments with regards to some item, claim, or public person, and medical scientists use 'pain scales' to gauge patients' subjective levels of discomfort and pain. Moreover, proportions and rates are ubiquitous across the social and life sciences (e.g., Douma & Weedon, 2019; Kubinec, 2020; P. Liu et al., 2020; Queiroz & Lemonte, 2021; Santos & Bolfarine, 2015). These and similar data would often have properties well-captured by the zero-one inflated beta distribution or the ordered beta distribution. We encourage researchers to carefully consider the processes that might have generated their data and, on this basis, fit several models of

¹⁴By 'ethnographic' we mean simply 'the processes and products of research that document what people know, feel, and do in a way that situates those phenomena at specific times in the history of individual lives' (Handwerker, 2001, p. 7).

681 varying complexity and functional form following an iterative (Bayesian) workflow of model
682 fitting, diagnosing, checking, comparison, tuning, and eventual inference (Gelman et al.,
683 2020). Such a workflow does not have as its end goal the selection of a one ‘best’ model;
684 rather, like peering at white light through a prism, the successes and failures of fitting many
685 different models often lead to insights easily overlooked by simply fitting one or a few.

Author contributions

T.B. and B.G.P. conceptualized the manuscript. B.G.P. managed data collection for the greater project from which we analyzed a subset. T.B. wrote the manuscript and conducted all analyses with critical inputs from B.G.P.

Acknowledgements

We thank Aaron D. Lightner and Matti Vuorre for comments on earlier drafts. We thank Aarhus University Research Foundation for financial support.

Data Availability

Data and code to reproduce this manuscript is available at: <https://github.com/tbendixen/freelist-tutorial>.

References

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Iannone, R. (2021). *Rmarkdown: Dynamic documents for r*. Retrieved from <https://github.com/rstudio/rmarkdown>
- Apicella, C. L. (2018). High levels of rule-bending in a minimally religious and largely egalitarian forager population. *Religion, Brain & Behavior*, 8(2), 133–148.
- Atkinson, Q. D. (2018). Religion and expanding the cooperative sphere in kastom and christian villages on Tanna, Vanuatu. *Religion, Brain & Behavior*, 8(2), 149–167.
- Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evolution and Human Behavior*, 32(1), 41–49.
- Atran, S., Medin, D., Ross, N., Lynch, E., Vapnarsky, V., Ek', E. U., ... Baran, M. (2002). Folkecology, Cultural Epidemiology, and the Spirit of the Commons: A Garden Experiment in the Maya Lowlands, 19912001. *Current Anthropology*, 43(3), 421–450. <https://doi.org/10.1086/339528>
- Barrett, H. C. (2020). Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution and Human Behavior*, 41(5), 445–453. <https://doi.org/10.1016/j.evolhumbehav.2020.05.006>
- Bendixen, T. (2019). Sense or non-sense? A critical discussion of a recent evolutionary-cognitive approach to “folk-economic beliefs.” *Evolution, Mind and Behaviour*, 17(1), 29–47. <https://doi.org/10.1556/2050.2019.00011>
- Bendixen, T. (2020). How cultural evolution can inform the science of science communication—and vice versa. *Humanities and Social Sciences Communications*, 7(1), 135. <https://doi.org/10.1057/s41599-020-00634-4>
- Bendixen, T., Apicella, C., Atkinson, Q., Cohen, E., Henrich, J., McNamara, R. A., ... Purzycki, B. G. (n.d.). Appealing to the minds of gods: A novel cultural evolutionary account of religious appeals and an empirical assessment using ethnographic data from eight diverse societies. <https://doi.org/10.31234/osf.io/tjn3e>

- Bendixen, T., Lightner, A. D., & Purzycki, B. G. (n.d.). The cultural evolution of religion and cooperation. In R. Kendal, J. Tehrani, & J. Kendal (Eds.), *Oxford handbook of cultural evolution*. Oxford University Press. <https://doi.org/10.31234/osf.io/fhscv>
- Bendixen, T., & Purzycki, B. G. (n.d.). Accounting for cross-cultural variation in the minds of gods. In B. G. Purzycki & T. Bendixen (Eds.), *The minds of gods: A cross-disciplinary survey*. Bloomsbury Publishing.
- Bernard, H. R. (2018). *Research methods in cultural anthropology. Qualitative and* (6th ed.).
- Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences*, *111*(47), 16784–16789. <https://doi.org/10.1073/pnas.1408701111>
- Brewer, D. D., Garrett, S. B., & Rinaldi, G. (2002). Free-listed items are effective cues for eliciting additional items in semantic domains. *Applied Cognitive Psychology*, *16*(3), 343–358. <https://doi.org/https://doi.org/10.1002/acp.797>
- Broesch, T., Crittenden, A., Beheim, B., Blackwell, A., Bunce, J., Colleran, H., . . . Mulder, M. (2020). Navigating cross-cultural research: Methodological and ethical considerations. *Proceedings. Biological Sciences*, *287*, 20201245. <https://doi.org/10.1098/rspb.2020.1245>
- Bunce, J. A., & McElreath, R. (2017). Interethnic Interaction, Strategic Bargaining Power, and the Dynamics of Cultural Norms: A Field Study in an Amazonian Population. *Human Nature*, *28*(4), 434–456. <https://doi.org/10.1007/s12110-017-9297-8>
- Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451. <https://doi.org/https://doi.org/10.1111/bmsp.12195>
- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).
- Chaves, L. da S., Nascimento, A. L. B. do, Albuquerque, U. P., Chaves, L. da S., Nascimento, A. L. B. do, & Albuquerque, U. P. (2019). What matters in free listing? A probabilistic interpretation of the salience index. *Acta Botanica Brasilica*, *33*(2), 360–369. <https://doi.org/10.1590/0102-33062018abb0330>
- Cohen, E., Baimel, A., & Purzycki, B. G. (2018). Religiosity and resource allocation in Marajó, Brazil. *Religion, Brain Behavior*, *8*(2), 168–184. <https://doi.org/10.1080/2153599X.2016.1267029>
- Cribari-Neto, F., & Santos, J. (2019). Inflated Kumaraswamy distributions. *Anais Da Academia Brasileira de Ciências*, *91*. <https://doi.org/10.1590/0001-3765201920180955>

- 769 D'Andrade, R. G. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.),
 770 *Cultural models in language and thought* (pp. 112–148). Cambridge University Press.
 771 <https://doi.org/10.1017/CBO9780511607660.006>
- 772 D'Andrade, R. G. (1995). *The Development of Cognitive Anthropology*. New York: Cam-
 773 bridge University Press.
- 774 Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A causal framework for cross-cultural
 775 generalizability. PsyArXiv. <https://doi.org/10.31234/osf.io/fqukp>
- 776 Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and
 777 evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology*
 778 *and Evolution*, 10(9), 1412–1430. <https://doi.org/10.1111/2041-210X.13234>
- 779 Dressler, W. W., Borges, C. D., Balieiro, M. C., & Santos, J. E. dos. (2005). Measuring
 780 Cultural Consonance: Examples with Special Reference to Measurement Theory in An-
 781 thropology. *Field Methods*, 17(4), 331–355. <https://doi.org/10.1177/1525822X05279899>
- 782 Erler, N. S., Rizopoulos, D., Rosmalen, J. van, Jaddoe, V. W. V., Franco, O. H., & Lesaffre,
 783 E. M. E. H. (2016). Dealing with missing covariates in epidemiologic studies: A compar-
 784 ison between multiple imputation and a full Bayesian approach. *Statistics in Medicine*,
 785 35(17), 2955–2974. <https://doi.org/10.1002/sim.6944>
- 786 Fremout, T., Gutiérrez-Miranda, C. E., Briers, S., Marcelo-Peña, J. L., Cueva-Ortiz, E.,
 787 Linares-Palomino, R., ... Thomas, E. (2021). The value of local ecological knowledge
 788 to guide tree species selection in tropical dry forest restoration. *Restoration Ecology*,
 789 29(4), e13347. <https://doi.org/https://doi.org/10.1111/rec.13347>
- 790 Furlow, C. A. (2003). Comparing indicators of knowledge within and between cultural
 791 domains. *Field Methods*, 15(1), 51–62.
- 792 Gabry, J., & Mahr, T. (2021). Bayesplot: Plotting for bayesian models. Retrieved from
 793 <https://mc-stan.org/bayesplot/>
- 794 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization
 795 in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in*
 796 *Society)*, 182(2), 389–402. <https://doi.org/10.1111/rssa.12378>
- 797 Ge, E., Chen, Y., Wu, J., & Mace, R. (2019). Large-scale cooperation driven by reputation,
 798 not fear of divine punishment. *Royal Society Open Science*, 6(8), 190991. <https://doi.org/10.1098/rsos.190991>
- 800 Gelman, A., Carlin, J. B., Stern, D., Hal S, & Rubin, D. B. (2013). *Bayesian data analysis*
 801 (3rd ed.). Chapman; Hall/CRC.
- 802 Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood
 803 in the context of the likelihood. *Entropy*, 19(10). <https://doi.org/10.3390/e19100555>
- 804 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ...
 805 Modrák, M. (2020). Bayesian Workflow. *arXiv:2011.01808 [Stat]*. Retrieved from <http://arxiv.org/abs/2011.01808>
- 806 <http://arxiv.org/abs/2011.01808>
- 807 Haaf, J. M., Merkle, E. C., & Rouder, J. N. (2020). Do items order? The psychology in
 808 IRT models. *Journal of Mathematical Psychology*, 98, 102398. <https://doi.org/10.1016/j.jmp.2020.102398>
- 809 j.jmp.2020.102398
- 810 Handwerker, W. P. (2001). *Quick ethnography*. Lanham: Altamira Press.
- 811 Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of*
 812 *Verbal Learning and Verbal Behavior*, 8(2), 176–184. [https://doi.org/https://doi.org/10.1016/S0022-5371\(69\)80058-7](https://doi.org/https://doi.org/10.1016/S0022-5371(69)80058-7)
- 813 10.1016/S0022-5371(69)80058-7

- 814 Jamieson-Lane, A., & Purzycki, B. G. (2016). *AnthroTools: Some custom tools for anthro-*
 815 *pology*.
- 816 Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). *Bayes Rules! An Introduction to Applied*
 817 *Bayesian Modeling*. Chapman; Hall/CRC Press.
- 818 Johnson, D. D. P. (2016). *God is watching you: How the fear of god makes us human*.
 819 Oxford University Press, USA.
- 820 Kay, M. (2022). *tidybayes: Tidy data and geoms for Bayesian models*. [https://doi.org/10.](https://doi.org/10.5281/zenodo.1308151)
 821 [5281/zenodo.1308151](https://doi.org/10.5281/zenodo.1308151)
- 822 Krosnick, J. A., & Presser, S. (2009). Question and Questionnaire Design. In J. D. Wright &
 823 P. V. Marsden (Eds.), *Handbook of Survey Research* (Second). San Diego, CA: Elsevier.
- 824 Kruschke, J. K. (2014). Doing bayesian data analysis: A tutorial with r, JAGS, and stan.
- 825 Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psycho-*
 826 *nomics Bulletin & Review*, 25(1), 155–177. <https://doi.org/10.3758/s13423-017-1272-1>
- 827 Kubinec, R. (2020). Ordered beta regression: A parsimonious, well-fitting model for survey
 828 sliders and visual analog scales. SocArXiv. <https://doi.org/10.31235/osf.io/2sx6y>
- 829 Kurz, A. S. (2021a). *Doing Bayesian data analysis in brms and the tidyverse* (version 0.4.0).
 830 Retrieved from <https://bookdown.org/content/3686/>
- 831 Kurz, A. S. (2021b). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second*
 832 *edition* (version 0.2.0). Retrieved from <https://bookdown.org/content/4857/>
- 833 Lang, M., Purzycki, B. G., Apicella, C. L., Atkinson, Q. D., Bolyanatz, A., Cohen, E.,
 834 ... others. (2019). Moralizing gods, impartiality and religious parochialism across 15
 835 societies. *Proceedings of the Royal Society B*, 286(1898), 20190202.
- 836 Levine, J., Muthukrishna, M., Chan, K., & Satterfield, T. (2015). Theories of the deep:
 837 Combining salience and network analyses to produce mental model visualizations of a
 838 coastal British Columbia food web. *Ecology and Society*, 20(4). [https://doi.org/10.](https://doi.org/10.5751/ES-08094-200442)
 839 [5751/ES-08094-200442](https://doi.org/10.5751/ES-08094-200442)
- 840 Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices
 841 based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9),
 842 1989–2001.
- 843 Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models:
 844 What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–
 845 348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- 846 Liu, F., & Eugenio, E. C. (2018). A review and comparison of bayesian and likelihood-based
 847 inferences in beta regression and zero-or-one-inflated beta regression. *Statistical Methods*
 848 *in Medical Research*, 27(4), 1024–1044. <https://doi.org/10.1177/0962280216650699>
- 849 Liu, F., & Kong, Y. (2015). Zoib: An R Package for Bayesian Inference for Beta Regression
 850 and Zero/One Inflated Beta Regression. *The R Journal*, 7(2), 34. [https://doi.org/10.](https://doi.org/10.32614/RJ-2015-019)
 851 [32614/RJ-2015-019](https://doi.org/10.32614/RJ-2015-019)
- 852 Liu, P., Yuen, K. C., Wu, L.-C., Tian, G.-L., & Li, T. (2020). Zero-one-inflated simplex
 853 regression models for the analysis of continuous proportion data. *Statistics and Its*
 854 *Interface*, 13(2), 193–208. <https://doi.org/10.4310/SII.2020.v13.n2.a5>
- 855 McElreath, R. (2020). *Statistical Rethinking: A Bayesian course with examples in R and*
 856 *Stan* (Second). CRC Press.
- 857 McNamara, R. A., & Henrich, J. (2018). Jesus vs. The ancestors: How specific religious
 858 beliefs shape prosociality on Yasawa Island, Fiji. *Religion, Brain & Behavior*, 8(2),

- 185–204.
- McNamara, R. A., Senanayake, R., Willard, A. K., & Henrich, J. (2021). God’s mind on morality. *Evolutionary Human Sciences*, 3, E6. <https://doi.org/10.1017/ehs.2021.1>
- McNamara, R. A., & Wertz, A. E. (2021). Early Plant Learning in Fiji. *Human Nature*, 32(1), 115–149. <https://doi.org/10.1007/s12110-021-09389-6>
- Menezes, A. F. B., Mazucheli, J., & Bourguignon, M. (2021). A parametric quantile regression approach for modelling zero-or-one inflated double bounded data. *Biometrical Journal*, 63(4), 841–858. <https://doi.org/10.1002/bimj.202000126>
- Mikkola, P., Martin, O. A., Chandramouli, S., Hartmann, M., Pla, O. A., Thomas, O., . . . Klami, A. (2021). Prior knowledge elicitation: The past, present, and future. Retrieved from <https://arxiv.org/abs/2112.01380>
- Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences*, 39.
- Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>
- Panter, A. T., Tanaka, J. S., & Wellens, T. R. (1992). The Psychometrics of Order Effects. In N. Schwarz & S. Sudman (Eds.), *Context Effects in Social and Psychological Research* (pp. 249–264). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-2848-6_17
- Purzycki, B. G. (2011). Tyvan *cher eezi* and the socioecological constraints of supernatural agents’ minds. *Religion, Brain & Behavior*, 1(1), 31–45.
- Purzycki, B. G. (2013). Toward a cognitive ecology of religious concepts: An example from the tyva republic. *Journal for the Cognitive Science of Religion*, 1(1), 99–120.
- Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., . . . Henrich, J. (2016a). Cross-cultural dataset for the evolution of religion and morality project. *Scientific Data*, 3, 160099. <https://doi.org/10.1038/sdata.2016.99>
- Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., . . . Henrich, J. (2016b). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, 530(7590), 327–330. <https://doi.org/10.1038/nature16980>
- Purzycki, B. G., & Bendixen, T. (2020). Examining Values, Virtues, and Tradition in the Tyva Republic with Free-List and Demographic Data. *The New Research in Tuva*, 4, 6–18.
- Purzycki, B. G., Henrich, J., Apicella, C., Atkinson, Q. D., Baimel, A., Cohen, E., . . . Norenzayan, A. (2018). The evolution of religion and morality: A synthesis of ethnographic and experimental evidence from eight societies. *Religion, Brain & Behavior*, 8(2), 101–132.
- Purzycki, B. G., & Holland, E. C. (2019). Buddha as a God: An Empirical Assessment. *Method & Theory in the Study of Religion*, 31(4-5), 347–375. <https://doi.org/10.1163/15700682-12341453>
- Purzycki, B. G., & Jamieson-Lane, A. (2016). AnthroTools: An R package for cross-cultural ethnographic data analysis. *Cross-Cultural Research*, 51(1), 51–74.
- Purzycki, B. G., & Kulundary, V. (2018). Buddhism, identity, and class: Fairness and favoritism in the tyva republic. *Religion, Brain & Behavior*, 8(2), 205–226.

- Purzycki, B. G., & McNamara, R. A. (2016). An ecological theory of gods' minds. In H. De Cruz & R. Nichols (Eds.), *Cognitive Science of Religion and Its Philosophical Implications* (pp. 143–167). New York: Continuum.
- Purzycki, B. G., Pisor, A., Apicella, C., Atkinson, Q. D., Cohen, E., Henrich, J., . . . Xygalatas, D. (2018). The cognitive and cultural foundations of moral behavior. *Evolution and Human Behavior*, 39(5), 490–501.
- Purzycki, B. G., Stagnaro, M. N., & Sasaki, J. (2020). Breaches of Trust Change the Content and Structure of Religious Appeals. *Journal for the Study of Religion, Nature and Culture*, 14(1), 71–94. <https://doi.org/10.1558/jsrnc.38786>
- Purzycki, B. G., Willard, A. K., Klocová, E. K., Apicella, C., Atkinson, Q., Bolyanatz, A., . . . Ross, C. T. (2022). The Moralization Bias of Gods' Minds: A Cross-Cultural Test. *Religion, Brain and Behavior*.
- Queiroz, F. F., & Lemonte, A. J. (2021). A broad class of zero-or-one inflated regression models for rates and proportions. *Canadian Journal of Statistics*, 49(2), 566–590. <https://doi.org/10.1002/cjs.11576>
- Quinlan, M. B. (2005). Considerations for collecting freelists in the field: Examples from ethobotany. *Field Methods*, 17(3), 219–234.
- Quinlan, M. B. (2017). The freelisting method. *Handbook of Research Methods in Health Social Sciences*. Singapore: Springer, 1–16.
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ragland, D. R. (1992). Dichotomizing Continuous Outcome Variables: Dependence of the Magnitude of Association and Statistical Power on the Cutpoint. *Epidemiology*, 3(5), 434–440. Retrieved from <https://www.jstor.org/stable/3702637>
- Rodríguez-Sánchez, F., Jackson, C. P., & Hutchins, S. D. (2022). *Grateful: Facilitate citation of r packages*. Retrieved from <https://github.com/Pakillo/grateful>
- Romney, A. K., & d'Andrade, R. G. (1964). Cognitive aspects of english kin terms. *American Anthropologist*, 66(3), 146–170.
- Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, 88(2), 313–338.
- Şahin, M. D. (2021). Effect of Item Order on Certain Psychometric Properties: A Demonstration on a Cyberloafing Scale. *Frontiers in Psychology*, 12, 154. <https://doi.org/10.3389/fpsyg.2021.590545>
- Santos, B., & Bolfarine, H. (2015). Bayesian analysis for zero-or-one inflated proportion data using quantile regression. *Journal of Statistical Computation and Simulation*, 85(17), 3579–3593. <https://doi.org/10.1080/00949655.2014.986733>
- Schloss, J. P., & Murray, M. J. (2011). Evolutionary accounts of belief in supernatural punishment: A critical review. *Religion, Brain & Behavior*, 1(1), 46–99.
- Simpson, A., Piazza, J., & Rios, K. (2016). Belief in divine moral authority: Validation of a shortened scale with implications for social attitudes and moral cognition. *Personality and Individual Differences*, 94, 256–265. <https://doi.org/10.1016/j.paid.2016.01.032>
- Singh, M., & Henrich, J. (2020). Why do religious leaders observe costly prohibitions? Examining taboos on Mentawai shamans. *Evolutionary Human Sciences*, 2. <https://doi.org/10.1017/ehs.2020.32>

- Singh, M., Kaptchuk, T. J., & Henrich, J. (2021). Small gods, rituals, and cooperation: The Mentawai water spirit Sikameinan. *Evolution and Human Behavior*, 42(1), 61–72.
- Smith, J. J. (1993). Using ANTHROPAC 3.5 and a Spreadsheet to Compute a Free-List Saliency Index. *CAM*, 5(3), 1–3. <https://doi.org/10.1177/1525822X9300500301>
- Smith, J. J., & Borgatti, S. P. (1997). Saliency Counts—And So Does Accuracy: Correcting and Updating a Measure for Free-List-Item Saliency. *Journal of Linguistic Anthropology*, 7(2), 208–209. <https://doi.org/10.1525/jlin.1997.7.2.208>
- Smith, J. J., Furbee, L., Maynard, K., Quick, S., & Ross, L. (1995). Saliency Counts: A Domain Analysis of English Color Terms. *Journal of Linguistic Anthropology*, 5(2), 203–216. <https://doi.org/10.1525/jlin.1995.5.2.203>
- Stan Development Team. (2021). RStan: The R interface to Stan. Retrieved from <https://mc-stan.org/>
- Turpin, H. D., & Stanford, M. (2020). Cognitively informed ethnography: Using mixed methods to capture the complexity of religious phenomena in two ecologically valid settings. *Journal for the Cognitive Science of Religion*, 0(0). <https://doi.org/10.1558/jcsr.38498>
- Ushey, K. (2022). *renv: Project environments*. Retrieved from <https://CRAN.R-project.org/package=renv>
- Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models. Retrieved from <https://mc-stan.org/loo/>
- 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>
- Vuorre, M. (2019). How to analyze visual analog (slider) scale data? Retrieved from <https://vuorre.netlify.app/post/2019/02/18/analyze-analog-scale-ratings-with-zero-one-inflated-beta-models/>
- Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1804), 20142556. <https://doi.org/10.1098/rspb.2014.2556>
- White, C. J. M., & Norenzayan, A. (2021). Karma and God: Convergent and divergent mental representations of supernatural norm enforcement. *Psychology of Religion and Spirituality*. <https://doi.org/10.1037/rel0000436>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Willard, A. K. (2018). Religion and prosocial behavior among the indo-fijians. *Religion, Brain & Behavior*, 8(2), 227–242. <https://doi.org/10.1080/2153599X.2016.1267032>
- Willard, A. K., Baimel, A., Turpin, H., Jong, J., & Whitehouse, H. (2020). Rewarding the good and punishing the bad: The role of karma and afterlife beliefs in shaping moral norms. *Evolution and Human Behavior*, 41(5), 385–396. <https://doi.org/10.1016/j.evolhumbehav.2020.07.001>

- 994 Xie, Y., Allaire, J. J., & Grolemond, G. (2018). *R markdown: The definitive guide*. Boca
995 Raton, Florida: Chapman; Hall/CRC. Retrieved from [https://bookdown.org/yihui/](https://bookdown.org/yihui/rmarkdown)
996 [rmarkdown](https://bookdown.org/yihui/rmarkdown)
- 997 Xygalatas, D., Kotherová, S., Maño, P., Kundt, R., Cigán, J., Klocová, E. K., & Lang, M.
998 (2018). Big gods in small places: The random allocation game in mauritius. *Religion,*
999 *Brain & Behavior*, 8(2), 243–261. <https://doi.org/10.1080/2153599X.2016.1267033>