Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution

Jeff S. Wesner and *Justin P.Z. Pomeranz*

University of South Dakota, Department of Biology, Vermillion, SD 57069

Jeff.Wesner@usd.edu

Keywords: *Bayesian, prior predictive distribution, modeling, GLMM, simulation*

Abstract

Bayesian data analysis is increasingly used in ecology, but prior specification remains focused on choosing non-informative priors (e.g., flat or vague priors). One barrier to choosing more informative priors is that priors must be specified on model parameters (e.g., intercepts, slopes, sigmas), but prior knowledge largely exists on the level of the response variable. This is particularly true for common models in ecology, like generalized linear mixed models, which may have a link function and dozens of parameters, each of which needs a prior distribution. We suggest that this difficulty can be overcome by simulating from the prior predictive distribution and visualizing the results on the scale of the response variable. In doing so, some common choices for non-informative priors on parameters can easily be seen to produce biologically impossible values of response variables. Such implications of prior choices are difficult to foresee without visualization. We demonstrate a workflow for prior selection using simulation and visualization with two ecological examples (predator-prey body sizes and spider responses to food competition). This approach is not new, but its adoption by ecologists will help to better incorporate prior information in ecological models, thereby maximizing one of the benefits of Bayesian data analysis.

Introduction

The distinguishing feature between Bayesian and non-Bayesian statistics is that Bayesian statistics treats unknown parameters as random variables governed by a probability distribution, while non-Bayesian statistics treats unknown parameters as fixed (Ellison and Dennis 2010, Hobbs and Hooten 2015). A common misconception is that only Bayesian statistics incorporates prior information. However, non-Bayesian methods can and often do incorporate prior information, either informally in the choices of likelihoods and model structures, or formally as penalized likelihood or hierarchical modeling (Hobbs and Hooten 2015, Morris et al. 2015).

While prior information is not unique to Bayesian models, it is required of them. For example, in a simple linear regression of the form $y \sim N(\alpha + \beta x, \sigma)$ intercept α, slope β, and error σ are unknown parameters that need a prior probability distribution. There are differing opinions and philosophies on the best practices for choosing priors (Lindley 1961, Edwards et al. 1963, Morris et al. 2015, Wolf et al. 2017, Lemoine 2019, Banner et al. 2020, Gelman et al. 2017). In ecology, a common practice is to assign so-called non-informative priors that allow Bayesian inference to proceed (i.e. produce a posterior distribution), but with limited influence of the priors (Gelman et al. 2013). The reasons for this are varied but are at least in part driven by a desire to avoid the appearance of subjectivity and/or a reliance on default priors in popular software (Gelman and Hennig 2017, Banner et al. 2020).

Regardless of the philosophy for choosing priors, it is important to understand what the priors actually encode in the model (Kennedy et al. 2019, Gabry et al. 2019). Even for seemingly simple and routine models, like logistic regression, it can be difficult to understand *a priori* how priors affect the model, because they must be assigned in the context of likelihood (Gelman et al. 2017). The result is that prior selection takes place on parameters that are often less intuitive to understand than the ultimate target of the analysis, such as the expected mean or raw y data (Kadane et al. 1980, Bedrick et al. 1996, James et al. 2010). In other words, ecologists that are interested in understanding the proportion of individuals that are infected with a virus (i.e., the outcome variable), need to assign priors on parameters that operate through a logit-link (in binomial regression) that is several steps removed from the measure of interest. This potentially leads to unintended mistakes in which a seemingly non-informative prior at the level of the individual parameter becomes informative on the outcome scale (Northrup and Gerber 2018, Banner et al. 2020, Gabry et al. 2019). While this type of mistake might be obvious to an experienced statistician with domain knowledge of the problem at hand, it is quite difficult to conceptualize for the rest of us (Kadane et al. 1980, Bedrick et al. 1996). This is particularly true for the types of models that are commonly used in ecology, such as generalized linear mixed models with interactions, which may have dozens of parameters, each of which

require a prior probability distribution (Bedrick et al. 1996, McElreath 2020).

We suggest that ecologists can address this problem using simulation from the prior predictive distribution and visualizing the implications of the priors on either the expected mean of the data (e.g., simulate regression lines or group means) (Kadane et al. 1980, Bedrick et al. 1996) or by simulating individual data points (Gabry et al. 2019). In this paper, we demonstrate how to use the prior predictive distribution to assign priors using two case studies with ecological data.

Prior Predictive Simulation

An attractive feature of the Bayesian approach is that the models are generative. This means that we can simulate potential data from the model so long as the parameters are assigned a proper probability distribution (Gelman et al. 2013). This feature is routinely used to check models after fitting the data using the posterior predictive distribution (Gelman et al. 2020), but it can also be used before seeing the data using the prior predictive distribution (Gabry et al. 2019). A simple model from a normal linear regression, using the syntax of (McElreath 2020), is:

$$y_i^{sims} \sim N(\mu_i^{sims}, \sigma^{sims})$$

$$\mu_i^{sims} = \alpha^{sims} + \beta^{sims} x_i$$

$$\alpha^{sims} \sim N(0, \sigma_\alpha)$$

$$\beta^{sims} \sim N(0, \sigma_\beta)$$

$$\sigma^{sims} \sim Exponential(\phi)$$

where y^{sims} are predicted values of the response variable, μ^{sims} are simulated values of the expected mean of the outcome, σ^{sims} are simulated errors as standard deviations, σ_α and σ_β are the standard deviations for the normal α and β priors, respectively, and ϕ is the rate parameter for the exponential distribution. We use the exponential distribution as a prior for standard deviations because it generates only positive values and allows for occasionally large deviations. For standard deviations (or variances), there are a number of alternatives prior distributions available (Gelman and others 2006, Gelman et al. 2013, McElreath 2020).

Simulations are made across each i value of the predictor variable(s) x. In this example, we assumed that x was known (e.g., if it were a gradient of planned Se additions to water before the experiment was conducted or a set of fixed treatments), but it could also be simulated.

With this general model, we do the following:

1) Draw N values from the prior distributions.
2) For each draw, solve the equation for each i value of x
3) Plot the result for either μ_i^{sims} or y_i^{sims}.
4) Use our domain knowledge (or another expert's) to assess whether the simulated values reflect prior knowledge.
5) If values do not reflect prior knowledge, change the prior distribution, likelihood, or both and repeat the simulation from step 1.
6) If values reflect prior knowledge, add the data and estimate the posterior distribution.

This amounts to a prior predictive check to satisfy the expectation that "simulations from the full Bayesian model... should be plausible data sets" (Kennedy et al. 2019). The simulation and visualization steps (1-3) are critical, simulated data sets are derived from the *joint distribution* of parameters. In other words, whether a model simulates plausible data data cannot be determined simply from looking at the individual priors or model formula, because their interpretation depends on the units of measurement (e.g., a N(0,1) prior means different things if y is measure in μm versus km) and on the range of prior expected values. We demonstrate this below with two motivating examples. All data and code are available at: https://github.com/jswesner/prior_predictive.

Motivating Examples

Example 1: Predator-Prey Body Sizes - Simple Linear Regression

Data

Understanding predator-prey interactions has long been a research interest of ecologists. Body size is related to a number of aspects that influence these interactions. For example, predators are often gape-limited, meaning that larger predators should be able to eat larger prey. The data set of (Brose et al. 2006) documents over 10,000 predator-prey interactions, including the mean mass of each.

Model

For this example, we examine the hypothesis that the mean prey body mass increases log-linearly with predator body mass using a simple linear model:

$$\log y_i \sim N(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \log x_i$$

$$\alpha \sim Normal(0, \sigma_\alpha)$$

$$\beta \sim Normal(0, \sigma_\beta)$$

$$\sigma \sim Exponential(\phi)$$

where log y_i is natural log transformed prey mass and log x_i is natural log transformed predator mass.

Priors

For the α and β priors, we need to specify a mean and standard deviation. In general, priors are made more informative by specifying a smaller standard deviation (or higher precision). Priors are "flattened" by giving them a large standard deviation (or smaller precision). For these priors, we assign a mean of 0 with a "non-informative" standard deviation of 1000 [$N(0, 1000)$]. There is nothing special about this prior, but it was a common default setting in earlier Bayesian software (usually specified as a precision rather than a standard deviation) and appears regularly in the literature, so we chose it as a representative starting point (Banner et al. 2020). Similarly, for the exponential distribution, smaller rates ϕ generate larger deviations, so we'll specify an initial ϕ of 0.00001. We chose this initial value by plotting 100 simulations from the exponential function in R under varying values of ϕ [e.g., *plot(rexp(100, 0.00001)*]. A value of 0.00001 generated an average deviance of ~1,000 with values up to ~5,000, indicating the possibility of producing extremely large values, particularly for log-transformed data.

After simulating from these initial priors, we re-fit the model with successfully tighter priors (Table 1) and compared the prior predictions to reference points (Mass of earth, a Blue Whale, a virus, and a Carbon-12 atom). The goal was to use these reference points to find a *joint prior distribution* that produced reasonable values of potential prey masses.

Results

Visualizing the implications of the priors on the scale of the response variable (log prey mass (g)) makes it clear that the wide "non-informative" priors make nonsense predictions (Figure 1a,b,d,e). In Figure 1a, all of the lines are impossibly steep, suggesting that predators could plausibly eat prey that are much larger than earth or much smaller than an atom. Even the seemingly narrower priors in Figure 1b suffer from the same problem, though the effect is clearly less severe. The narrowest priors (Figure 1c) produce more reasonable predictions, though they are still quite vague, with positive probability that large and small predators could eat prey that are orders of magnitude larger than an adult Blue Whale.

The conditional distributions in Figure 1d,c,e tell a similar story. They show the prior prediction of the mass

of an individual prey that an averaged sized predator would eat. As before, both the widest and narrow priors (Figure 1d,e) indicate that most (Figure 1d) or some (Figure 1e) of the prior probability is placed on prey sizes that exceed earth or are smaller than an atom. More worrisome, in the model with the widest priors, the average prey size is predicted to be 418 log grams. That exponentiates to a prey that weighs $3.4x10^{181}$ grams, which is 6 times larger than the mass of earth. In other words, these seemingly non-informative priors on the model parameters become informative on the outcome scale of a derived quantity (Hobbs and Hooten 2015, Gabry et al. 2019). By contrast, the mean prey sizes in the narrow and narrowest models are 1.1 and 0.7 log grams, respectively. That converts to 3 or 2 grams, larger than a dragonfly but smaller than a frog.

After settling on the priors used in Figure 1c,d, we fit the model and plotted the prior against the posterior distribution. While the priors appeared somewhat informative relative to the priors with larger values of σ and smaller values of ϕ (Figure 1a-c), it is clear that they were still relatively weak compared to the posterior distribution (Figure 2a,b).

Caveats

In a real analysis, there are some other steps we could have taken to generate a more realistic prior distribution before fitting the model to data. First, centering the intercept at 0 seemed like a reasonable first approximation, but we know from the literature that predators are generally larger than their prey. Therefore, it might make sense to alter the mean of the intercept prior to a value below zero, perhaps using an average predator/prey mass comparison from the literature. Similarly, the fact that larger predators tend to eat larger prey is well-known, so the prior on the slope β could be changed to a non-zero mean. One option would be to restrict the slope to only positive values, but this would not reflect our prior knowledge that predator body size is still a noisy predictor of prey body size (e.g., whales eat prey that are orders of magnitude smaller than they are).

Part of the uncertainty in prior selection can also be minimized by standardizing predictors (McElreath 2020). This changes the scale of each predictor so that the interpretation of its associated parameter is in units of standard deviation. In other words, a value of 2.3 for β would indicate that y increases by 2.3 for every standard deviation increase in x. Standardizing the predictors makes them unitless, thereby removing problems that can arise by mistaking cm for m or ha for acres. It also limits the expected prior values (e.g., N(0, 10) is extremely vague on a standardized predictor, but might be informative on a non-standardized predictor), but at the cost of less intuitive interpretation.

Example 2: Spider Abundance - Generalized Linear Mixed Model

Data

This data set comes from (Warmbold and Wesner 2018), who studied how terrestrial spiders responded to different combinations of freshwater fish using fish enclosure cages in a backwater of the Missouri River, USA. The hypothesized mechanism was that fish would reduce the emergence of adult aquatic insects by eating the insects, causing a reduction in terrestrial spiders that feed on the adult forms of those insects. The original experiment contained six treatments. Here, we present a simplified version comparing spider abundance above three treatments that contain either Smallmouth Buffalo (*Ictiobus bubalus*), Green Sunfish (*Lepomis cyanellus*), or a fishless control. Each treatment had four replicates for a total of 12 cages (each $2.3m^2$). The number of occupied spider webs above each cage were counted on four dates over the 29-day experiment.

Model

We fit a generalized linear mixed model with a Poisson likelihood, since the response y is a non-negative integer (i.e. number of spiders counted above a cage on each date). The predictor variables were date, treatment, and a date x treatment interaction. Since each replicate cage was sampled four times, we included a random intercept for cages. Describing the model as having two main effects and an interaction is deceptively simple. In reality, the model has 13 parameters that that require a prior specification: 12 "fixed" effects that indicate all combinations of each date x treatment, plus a "random" effect on the intercept:

$$y_i \sim Poisson(\lambda_i)$$

$$\log \lambda_i = \alpha + \alpha_{[cage]} +$$

$$\beta_1 x_{trt_i=fishless} + \beta_2 x_{trt_i=green} + \beta_3 x_{date_i=2} + \beta_4 x_{date_i=3} +$$

$$\beta_5 x_{date4} + \beta_6 x_{trt_i=fishless:date_i=2} + \beta_7 x_{trt_i=green:date_i=2}$$

+

$$\beta_8 x_{trt_i=fishless:date_i=3} + \beta_9 x_{trt_i=green:date_i=3} + beta_{10} x_{trt_i=fishless:date_i=4}$$

+

$$\beta_{11} x_{trt_i=green:date_i=4}$$

$$\alpha \sim Normal(0, \sigma_\alpha)$$

$$\alpha_{[cage]} \sim N(0, \sigma_{cage})$$

$$\beta_{1...11} \sim Normal(0, \sigma_{\beta_{1...11}})$$

$$\sigma_{cage} \sim Exponential(\phi)$$

where each y_i is described by a Poisson distribution with mean λ_i. Because the likelihood is not Gaussian, we specify a log link log λ_i so that the mean can be estimated as a linear function of predictors. This also ensures that the mean will be a positive number, so that we do not fit a model that predicts negative spider abundance. In this model, the intercept α represents the predicted log mean number of spiders in the treatment with Smallmouth Buffalo on the first sample date. The choice of reference treatment is arbitrary. Choosing Smallmouth Buffalo and the first date as the intercept is the default choice in R (R Core Team 2020) simply because the treatment is coded first alphabetically ("buffalo") and first numerically ("2015-06-08").

Priors

As before, we simulated outcomes under three model scenarios, each with different priors (Table 1). Another complication in this model is the log-link, which changes the biological interpretation of the priors. With a log-link the individual model parameters are now even less intuitive than they would be under a Gaussian likelihood (Bedrick et al. 1996). Under a Gaussian likelihood, a $\beta_1 x_{trt_i=fishless}$ value of 1.5 would indicate that the fishless treatment on 2020-06-08 contains 1.5 more spiders on average than the Smallmouth Buffalo treatment on the same date. With a Poisson likelihood and log-link, the same value first needs to be exponentiated $exp(1.5) = 4.5$ and then interpreted as a multiplier. Thus, a value of 1.5 for the parameter indicates that the fishless treatment contains 4.5 *times* more spiders than the Smallmouth Buffalo treatment on the first sample date. A value of 10 results in 22,026 times more spiders. This is an example of the principle that the prior can only be understood in the context of the likelihood (Gelman et al. 2017).

With practice, it becomes easier to conceptualize the interpretation of priors through a link function, but it is still important to visualize their implications on the outcome scale. For example, the intercept in this model predicts log λ_i. Yet the outcome we want to know about is not the log mean number of spiders, but the raw mean number of spiders λ_i or the total number of spiders $Poisson(\lambda_i)$ (Figure 3). Thus, moving from the prior on α to the inference we seek involves at least one or two transformations. Visualization of this provides an essential aid for understanding our prior implications (McElreath 2020).

Results

If all we knew was that spiders were counted above $2.32m^2$ cages but we did not know anything else about the experiment (i.e. the ecosystem, the question, the spider taxa), then we could still use the prior predictive

distribution to make better informative priors. The widest priors place substantial probabilities on values of >100,000 spiders per cage *on average* (Figure 4a), and include a small number of predictions on the final sample date with more than 100 million spiders (Figure 4c). We looked up the range of spider masses (~0.0005 to 170 grams). If we assume our spiders are relatively small, say 0.01 grams, then 100 million spiders would equal 30 tons of spiders. This is approximately equal to the mass of ~6 adult hippopotamus's (each ~4 tons).

However, in this case we do have valuable prior information. In a previous study using the same cages in the same backwater, (Warmbold 2016) counted between 0 and 2 spiders per cage. The present experiment had a slightly different design, in which a small rope was added to the center of each cage to increase the area of attachment (and presumably the number of spiders that would colonize) (Warmbold and Wesner 2018). If we assume that the added rope will double the number of spiders that could colonize, then it seems reasonable to expect at least 4 spiders per cage to colonize. There is obvious error associated with this, since the experiment was conducted in a different year and a different month. For that reason, we chose the moderate prior (Figure 4b,d) to use in the final model. It places most of the prior probability on values between ~1 to 100 spiders, but also allows for some extreme possibilities of >1000 spiders per cage (Figure 4d). The more restrictive priors are also reasonable, placing most of the prior probability between ~1 to 10 spiders, while allowing for up to ~100 spiders in extreme cases (Figure 4c,e).

Figure 4 shows the results after fitting the model to data. Spider counts ranged from 0 to 5 spiders per cage (Figure 4a), resulting in mean spider densities of ~1 to 4 spiders among the date x treatment combinations with far less uncertainty than specified in the prior (Figure 4a). Simulating from the prior and posterior predictive distributions shows the model predictions for the number of spiders we might expect at a new cage (i.e. a cage from another place or time). Before seeing the data, the model suggested reasonable probabilities of collecting 10 to >100 spiders. After seeing the data, the model suggests that finding ~10 or more spiders would be surprising (Figure 4b).

Caveats

Each of the 11 β's was assigned an independent prior. An alternative approach would be to assign β priors from a multivariate normal distribution (Hobbs and Hooten 2015). In addition, the likelihood assumes that the variance is equal to the mean. An alternative likelihood, such as a negative binomial, would allow us to model variances independently.

Discussion

Bayesian statistics is increasingly used by ecologists (Ellison 2004, McCarthy and Masters 2005, Hooten and

Hobbs 2015, Touchon and McCoy 2016), yet the preponderance of studies continue to rely on diffuse and/or default priors (Lemoine 2019, Banner et al. 2020). Using two case studies with a linear regression and a generalized linear mixed model - two common types of models in ecology (Touchon and McCoy 2016) - we demonstrated how visualization on the scale of the outcome can improve our choices of priors on individual parameters in a Bayesian analysis. From our own experience teaching Bayesian statistics to graduate students (JSW) and the experiences of others (James et al. 2010, Gabry et al. 2019), we suspect that this approach will help to remove confusion or anxiety over choosing more informative priors by aligning the choices more closely to the domain expertise of the users (Bedrick et al. 1996, James et al. 2010).

Choosing priors based on their implications on the outcome scale is not new. (Kadane et al. 1980) described a similar approach with Gaussian linear regressions to elicit prior information from experts, and (Bedrick et al. 1996) expanded it to generalized linear models. More recently, (Gabry et al. 2019) used it in a model with random effects to measure global air quality. (Kennedy et al. 2019) used a similar approach for models in cognitive science. A primary difference between the earlier and later uses of prior predictive simulation is the improvement in visualization techniques (Gabry et al. 2019), which makes it easier evaluate prior choices on a visual *distribution* of outcome measures, rather than only point estimates.

Assessing and visualizing priors on the outcome scale of a model makes clear what many current Bayesian approaches emphasize: it almost never the case that we have absolutely zero prior information (Hobbs and Hooten 2015, Lemoine 2019, Banner et al. 2020). For example, it does not take expertise in ecology or in predator-prey interactions to know that predators cannot eat prey larger than earth, yet this type of impossible prior belief is exactly what many Bayesian models encode with non-informative priors. It *does* take ecological expertise to know whether it is more probable for predators to eat prey that are 2 times larger or 2 times smaller, or whether the log-linear model should have a different functional form (e.g., non-linear). Critiquing priors in this way would, we argue, lead to a much better use of Bayesian methods than current practices that focus on finding the least informative prior (Lemoine 2019, Banner et al. 2020). Even for models with more abstract outcomes than body size (e.g., gene methylation, stoichiometric ratios, pupation rates of a new insect species), it is almost always the case that ecologists have some sense of what reasonable measures might look like. After all, it would be impossible to do any sort of study without first knowing what we will measure. Prior expectations of those measures come either from prior experience, the literature, or most often, both.

Visualizing simulations from the prior predictive distribution represents one aspect of the overall Bayesian modeling workflow (Kennedy et al. 2019, Gelman et al. 2020, Schad et al. 2020, Gabry et al. 2019). Like any approach to data analysis, the Bayesian workflow involves iteratively checking assumptions and implications

of our model, from data collection and model design to prior choices and model inference (Hooten and Hobbs 2015, Gelman et al. 2020). Traditionally, the role of priors in this workflow has focused on choosing the least informative priors possible, leading to a large body of theoretical and applied literature on development of non-informative priors, such as Jeffrey's, Horseshoe, or flat priors (Hobbs and Hooten 2015). When prior criticism is used, it is usually done after the model is fit with prior sensitivity analyses and/or plots of prior versus posterior parameters (Korner-Nievergelt et al. 2015). The approach we demonstrate does not obviate the need for these techniques in any sense. Rather, it adopts the approaches that are generally reserved for exploring the implications of the posterior distribution and applies them to the prior distribution. In doing so, it helps to lessen the impact of poor prior distributions later in the analysis workflow.

In ecology, the most closely related application of the approach we describe is for eliciting prior information from a panel of experts (James et al. 2010). However, external elicitation is not practical for most ecological studies, because the data analyst is often also the domain expert (Ellison and Dennis 2010). In other words, most statistical analysis in ecology is done by people (such as us) that are trained in disciplines other than statistics (Touchon and McCoy 2016). As a result, Bayesian analysis in ecology has traditionally been limited to ecologists with advanced statistical and computing capabilities. This is in part because Bayesian analysis is not included by default in popular statistical software, such as R (R Core Team 2020), and also because of the large computing time needed to run Bayesian models relative to frequentist or maximum likelihood approaches. Yet recent improvements in both the MCMC algorithms (Gelman et al. 2015) and the packages used to fit models appear likely to continue the trend of ecologists using Bayesian statistics. For example, with the *brms* package in R (Burkner 2017), this frequentist linear regression - *lm(y ~ x, data = data)* - becomes this Bayesian regression by changing two letters - *brm(y ~ x, data = data)*. This represents the simplest of cases (priors can and should be specified in the *brm()* model), but demonstrates the ease with which fitting Bayesian models is now possible.

An added benefit to choosing more informative priors is that it reduces the computational time needed to fit models, because it limits the parameter space that an MCMC algorithm needs to explore. We do not want our MCMC sampler to propose a parameter value that implies that a fish could eat a prey item larger than a whale (or larger than earth). In the relatively simple models we used here, the computational improvements are likely minimal. But ecologists are using increasingly sophisticated models (Touchon and McCoy 2016), for which the improvements in computational efficiency are likely to be notable. An irony in this improvement is that it contradicts a common justification of using non-informative priors to "let the data speak for themselves". In a model with such priors, much of the "speaking" is done by the priors in the sense of sampling parameter spaces that are incompatible with reasonable data. More importantly, as shown

by the first analysis here and by (Gabry et al. 2019), non-informative priors on parameters can become informative for quantities of interest (e.g., average prey sizes that are larger than earth). To rearrange the statement, data can only speak for themselves if the microphone is properly tuned.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1837233. JSW thanks the students in his graduate Bayesian class for asking challenging questions. He hopes this manuscript provides a long overdue answer. We thank Brose et al. (2006) for making their data publicly available.

References

Banner, K. M., K. M. Irvine, and T. J. Rodhouse. 2020. The use of Bayesian priors in Ecology: The good, the bad and the not great. Methods in Ecology and Evolution 11:882–889.

Bedrick, E. J., R. Christensen, and W. Johnson. 1996. A new perspective on priors for generalized linear models. Journal of the American Statistical Association 91:1450–1460.

Brose, U., T. Jonsson, E. L. Berlow, P. Warren, C. Banasek-Richter, L.-F. Bersier, J. L. Blanchard, T. Brey, S. R. Carpenter, M.-F. C. Blandenier, and others. 2006. Consumer–resource body-size relationships in natural food webs. Ecology 87:2411–2417.

Burkner, P.-C. 2017. Brms: An R Package for Bayesian Generalized Linear Mixed Models using Stan:22.

Edwards, W., H. Lindman, and L. J. Savage. 1963. Bayesian statistical inference for psychological research. Psychological review 70:193.

Ellison, A. M. 2004. Bayesian inference in ecology. Ecology letters 7:509–520.

Ellison, A. M., and B. Dennis. 2010. Paths to statistical fluency for ecologists. Frontiers in Ecology and the Environment 8:362–370.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. 2019. Visualization in Bayesian workflow. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182:389–402.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis. CRC press.

Gelman, A., and C. Hennig. 2017. Beyond objective and subjective in statistics (with discussion). Journal of the Royal Statistical Society, Series A 180:967–1033.

Gelman, A., D. Lee, and J. Guo. 2015. Stan: A probabilistic programming language for Bayesian inference and optimization. Journal of Educational and Behavioral Statistics 40:530–543.

Gelman, A., and others. 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian analysis 1:515–534.

Gelman, A., D. Simpson, and M. Betancourt. 2017. The Prior Can Often Only Be Understood in the Context of the Likelihood. Entropy 19:555.

Gelman, A., A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. 2020. Bayesian workflow. arXiv preprint arXiv:2011.01808.

Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: A statistical primer for ecologists. Princeton University Press.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85:3–28.

James, A., S. L. Choy, and K. Mengersen. 2010. Elicitator: An expert elicitation tool for regression in ecology. Environmental Modelling & Software 25:129–145.

Kadane, J. B., J. M. Dickey, R. L. Winkler, W. S. Smith, and S. C. Peters. 1980. Interactive elicitation of opinion for a normal linear model. Journal of the American Statistical Association 75:845–854.

Kennedy, L., D. Simpson, and A. Gelman. 2019. The Experiment is just as Important as the Likelihood in Understanding the Prior: A Cautionary Note on Robust Cognitive Modeling. Computational Brain & Behavior 2:210–217.

Korner-Nievergelt, F., T. Roth, S. Von Felten, J. Guélat, B. Almasi, and P. Korner-Nievergelt. 2015. Bayesian data analysis in ecology using linear models with R, BUGS, and Stan. Academic Press.

Lemoine, N. P. 2019. Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. Oikos 128:912–928.

Lindley, D. V. 1961. The use of prior probability distributions in statistical inference and decision. Pages 453–468 *in* Proc. 4th Berkeley Symp. On Math. Stat. And Prob.

McCarthy, M. A., and P. Masters. 2005. Profiting from prior information in Bayesian analyses of ecological data. Journal of Applied Ecology:1012–1019.

McElreath, R. 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.

Morris, W. K., P. A. Vesk, M. A. McCarthy, S. Bunyavejchewin, and P. J. Baker. 2015. The neglected tool in the Bayesian ecologist's shed: A case study testing informative priors' effect on model accuracy. Ecology and Evolution 5:102–108.

Northrup, J. M., and B. D. Gerber. 2018. A comment on priors for Bayesian occupancy models. PloS one 13:e0192819.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schad, D. J., M. Betancourt, and S. Vasishth. 2020. Toward a principled Bayesian workflow in cognitive science. Psychological methods.

Touchon, J. C., and M. W. McCoy. 2016. The mismatch between current statistical practice and doctoral training in ecology. Ecosphere 7:e01394.

Warmbold, J. 2016. Effects of fish on aquatic and terrestrial ecosystems. University of South Dakota.

Warmbold, J. W., and J. S. Wesner. 2018. Predator foraging strategy mediates the effects of predators on local and emigrating prey. Oikos 127:579–589.

Wolf, C., M. Novak, and A. I. Gitelman. 2017. Bayesian characterization of uncertainty in species interaction strengths. Oecologia 184:327–339.

Tables

Table 1: Priors used for the two models. Distributions are either Gaussian with a mean and standard deviation [N(mu, sigma)] or exponential [Exp(rate)].

	Model 1: Predator-Prey			Model 2: Spiders		
Parameter	Wide	Narrow	Narrowest	Wide	Narrow	Narrowest
Alpha	N(0,1000)	N(0,10)	N(0,1)	N(0,10)	N(0,1)	N(0,0.1)
Beta(s)	N(0,1000)	N(0,10)	N(0,1)	N(0,10)	N(0,1)	N(0,0.1)
Sigma	Exp(0.00001)	Exp(0.01)	Exp(0.1)			
Sigma_alpha				Exp(0.1)	Exp(1)	Exp(2)
Sigma_cage				Exp(0.1)	Exp(1)	Exp(2)

Figures

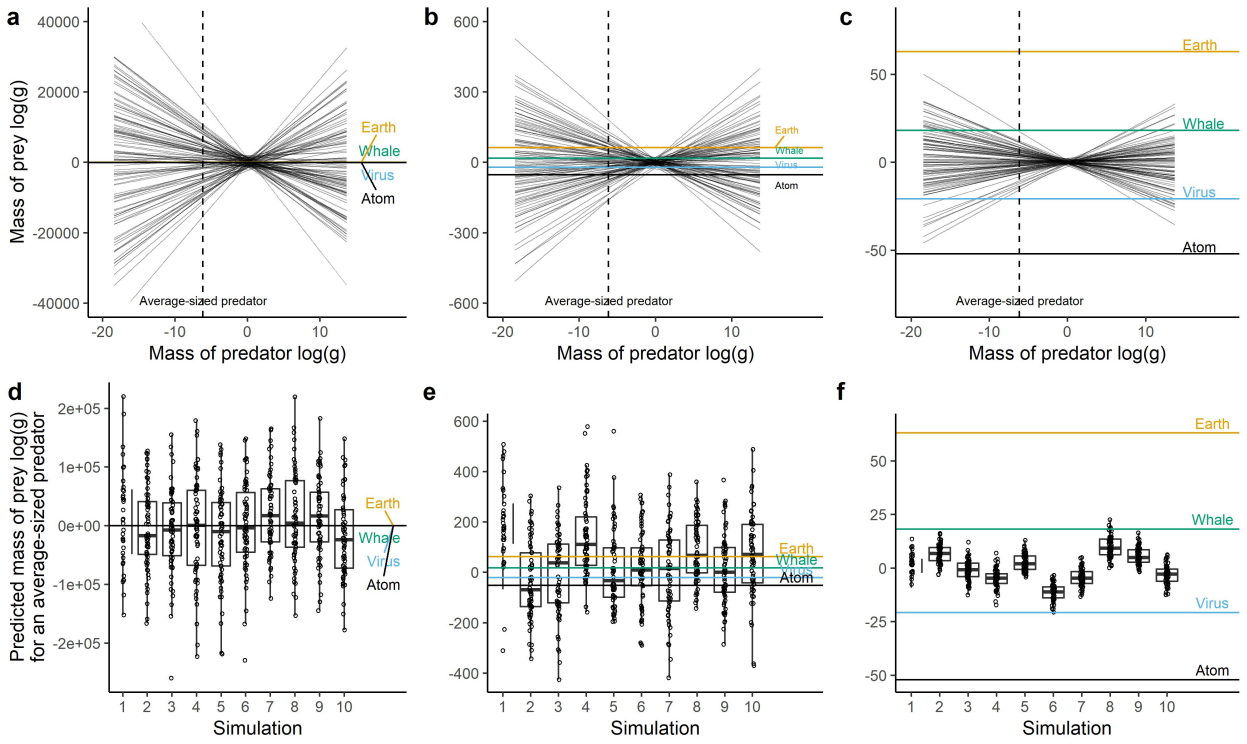

Figure 1: Prior predictive simulations showing the implications of the priors on predictions of log prey mass. Top row: 100 simulated regression lines from the prior predictive distributions of three models with either a) wide priors ($\sigma_{\alpha/\beta} = 1,000, \phi = 0.00001$), b) narrower priors ($\sigma_{\alpha/\beta} = 10, \phi = 0.01$), or c) the narrowest priors ($\sigma_{\alpha/\beta} = 1, \phi = 0.1$). Bottom row: 100 simulations of the predicted masses of prey for a median sized predator (-6 log(g)). Simulations come from the same priors as described above as d) wide priors e) narrower priors, and f) the narrowest priors.

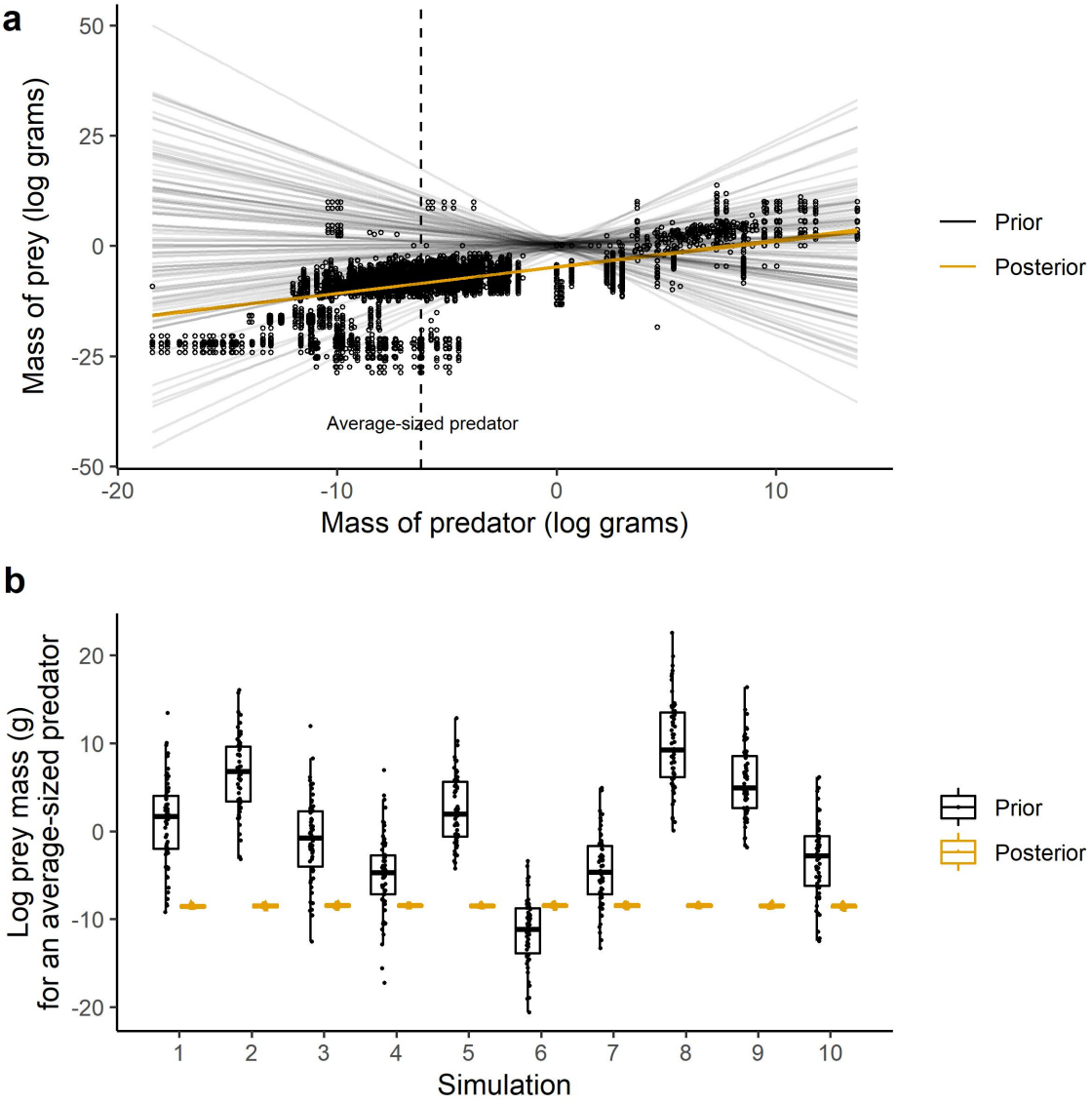

Figure 2: Comparison of the posterior distribution (orange) and the prior distribution (prior) for a) the regression lines and b) the conditional prediction of an average size of prey for an average sized predator. Each line is one of 100 simulations from either the prior or posterior distribution. The dots are the raw data. The prior is taken from the narrowest prior in Figure 1c. Even though this prior appeared to be somewhat informative relative to the other priors, it is clear that a large amount was learned from the data, as evidenced by the difference between the prior and posterior distributions in both panels.

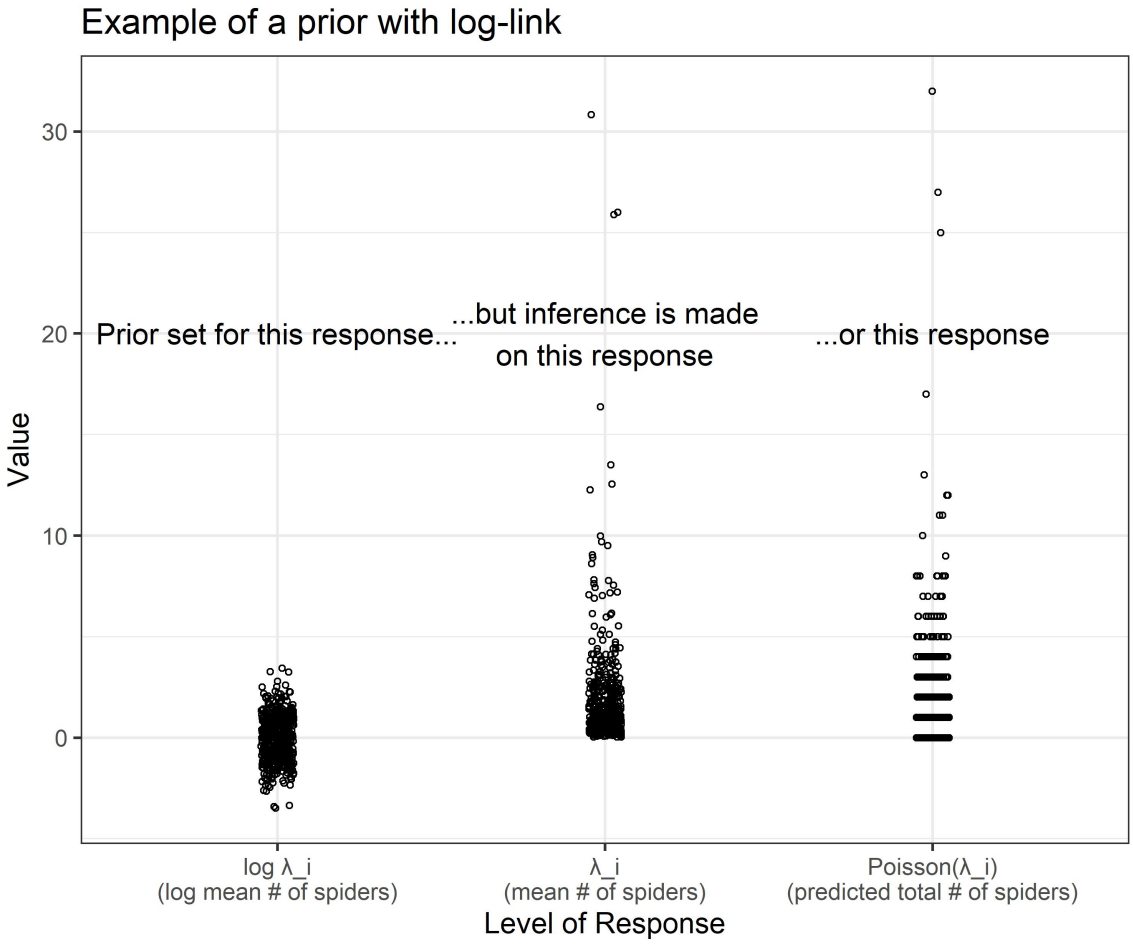

Figure 3: Different levels of inference from the same generalized linear mixed model for spider abundance. Simulations (n = 100) of either the log mean number of spiders (left), the mean number of spiders (middle), or the total number of spiders (right). Simulations come from a N(0,10) prior for the intercept α, which is specified based on log mean number of spiders, but is easier to assess on the raw mean number of spiders or the total number of spiders.

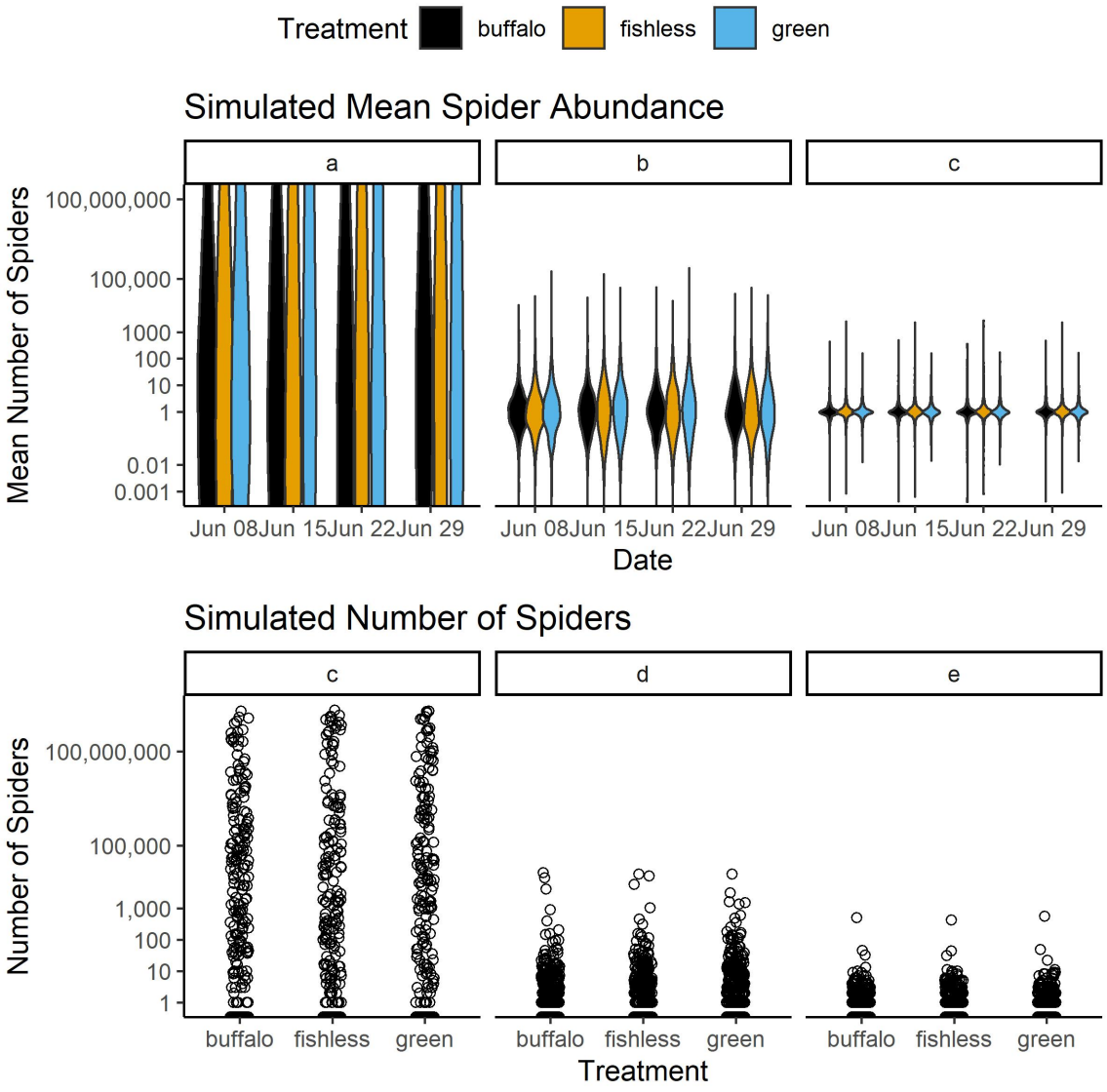

Figure 4: Prior predictive simulations showing the implications of the priors on spider densities above mesocosm cages. Top row: Prior predictive distribution of the number of the mean number of spiders above treatments with either Smallmouth Buffalo, no fish, or Green Sunfish. a) wide priors ($\sigma_{\alpha/\beta} = 10, \phi = 0.1$), b) narrower priors ($\sigma_{\alpha/\beta} = 1, \phi = 1$, or c) the narrowest priors ($\sigma_{\alpha/\beta} = 0.1, \phi = 2$). Bottom row: 500 simulations from the prior predictive distribution of the total number of spiders expected for a new cage. Simulations come from the same priors as described above as d) wide priors, e) narrower priors, and f) the narrowest priors. To improve visualation, the y-axis for a) is clipped at 0.001 and 1e9.

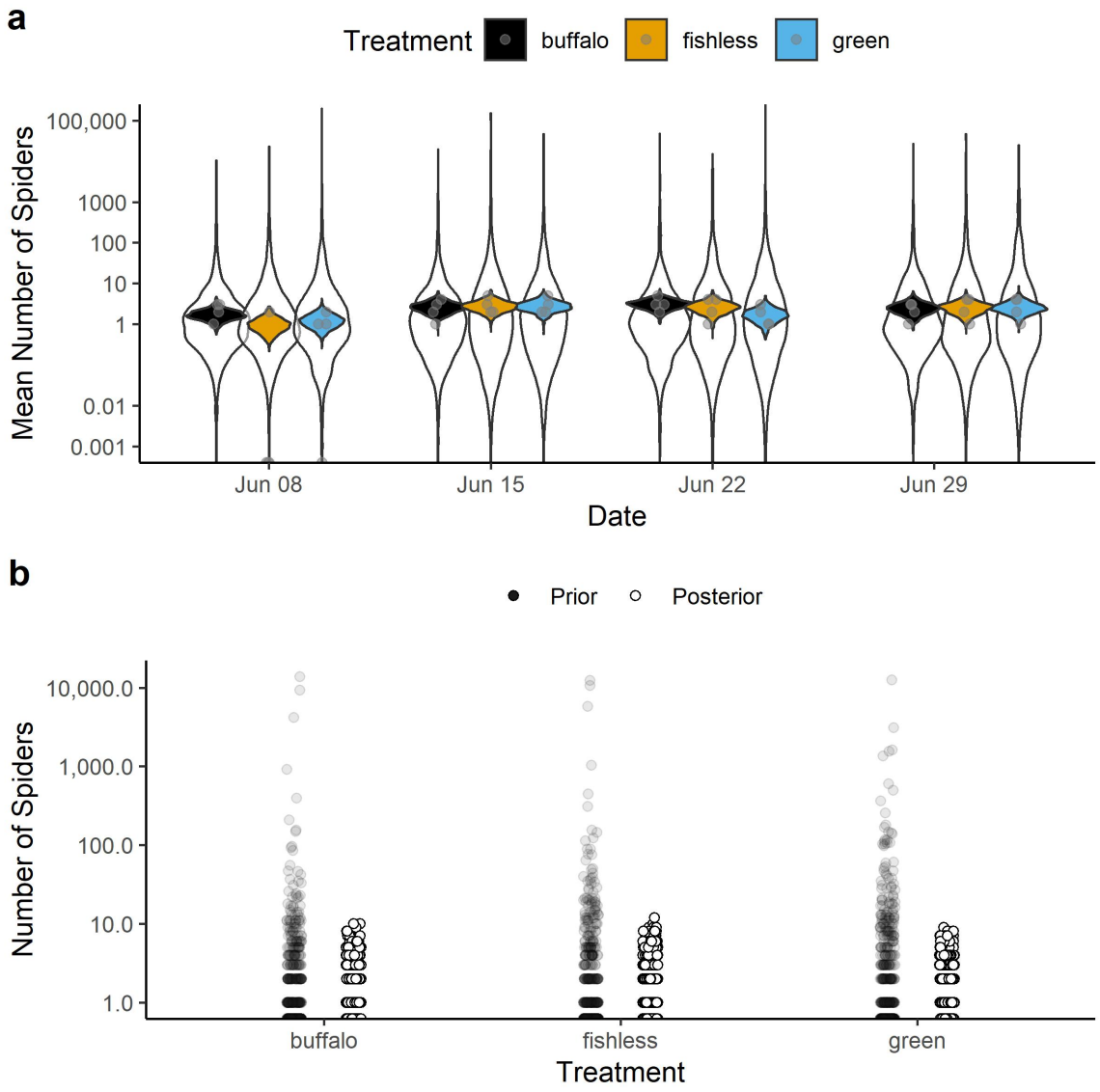

Figure 5: Comparison of the prior and posterior distributions for a) mean number of spiders and b) the conditional prediction of the number of spiders predicted for a new cage from each date x treatment combination. Each violin plot in (a) shows either the prior (white) or posterior (color) distribution with dots as raw data. Each dot in (b) is a simulation (n = 500) of the total number of spiders predicted for a single new cage in each date x treatment combination. The prior is taken from the narrow prior in Figure 4b. It is clear that a large amount was learned from the data, as evidenced by the difference between the prior and posterior distributions in both panels.