Prior Setting in Practice: Strategies and Rationales Used in Choosing Prior Distributions for Bayesian Analysis

Abhraneel Sarma
Northwestern University
abhraneel@u.northwestern.edu

Matthew Kay
University of Michigan
mjskay@umich.edu

ABSTRACT

Bayesian statistical analysis is steadily growing in popularity and use. Choosing priors is an integral part of Bayesian inference. While there exist extensive normative recommendations for prior setting, little is known about how priors are chosen in practice. We conducted a survey (N = 50) and interviews (N = 9) where we used interactive visualizations to elicit prior distributions from researchers experienced with Bayesian statistics and asked them for rationales for those priors. We found that participants' experience and philosophy influence how much and what information they are willing to incorporate into their priors, manifesting as different levels of informativeness and skepticism. We also identified three broad strategies participants use to set their priors: centrality matching, interval matching, and visual probability mass allocation. We discovered that participants' understanding of the notion of "weakly informative priors"—a commonly-recommended normative approach to prior setting—manifests very differently across participants. Our results have implications both for how to develop prior setting recommendations and how to design tools to elicit priors in Bayesian analysis.

Author Keywords

Bayesian inference; prior distributions; descriptive analysis

CCS Concepts

•**Human-centered computing** → *Visualization;*

INTRODUCTION

Bayesian statistical analysis has gained attention in recent years as an alternative to the traditional frequentist, or null hypothesis significance testing (NHST) approaches. This is partly because traditionally computationally expensive analyses, such as Markov Chain Monte Carlo (MCMC) sampling, have become more accessible: the increase in computational power now allow users to run such analyses using their personal computers. A growing number of modelling languages and software packages such as Stan [4], brms [3], rstanarm [40], and JAGs [43] also support Bayesian statistical analysis using notations that are closer to the way that statistical models are written mathematically (sometimes called *probabilistic programming languages*), allowing more users of varying skill levels to implement Bayesian analysis.

Another factor could be the growing calls for a focus on *estimation* instead of hypothesis testing [7, 37], as the Bayesian approach allows us to interpret results in terms of probabilities of particular effect sizes (conditional on prior knowledge), which traditional approaches do not [8, 24]. Further, it provides a principled mechanism for researchers to build on previous research by incorporating relevant prior information and domain knowledge into the analysis. This can be crucial in studies with small sample sizes, as prior information can be used to *regularize* results, reducing the chances of obtaining unrealistically large effect sizes due to chance—a phenomenon common in individual studies with relatively small sample sizes [27]. In HCI, small sample studies are abundant, and researchers have advocated the adoption of Bayesian methods as it allows for "more principled conclusions from small-n studies" [31].

Phelan et al.'s [42] work shows that certain technological interventions might help introduce users familiar with frequentist statistics to Bayesian methods. However, in their study, participants had difficulty grasping the concept of priors and specifying prior distributions. Though they mention that the prior distributions chosen by their participants were "reasonable", what constitutes a reasonable prior is difficult to define. The promised benefits of Bayesian analysis are partly contingent on specifying good priors, especially in smaller samples.

Since the choice of reasonable priors is dependent on factors such as the design of the experiment, data collection method, and the statistical model used for data analysis, specifying priors is an inherently a difficult task, especially for novices—Interactive prior elicitation interfaces could encode normative guidance in prior selection to better support this task. Existing literature offers such normative guidance, but little is known about how people with expertise in Bayesian statistical modelling choose priors in practice. Such descriptive knowledge could inform the development of interfaces for prior-setting designed for both experts and novices alike.

As a probe into existing prior-setting practice, we conducted two studies, an online survey and follow-up interviews. In both the studies, we presented participants with a common statistical model and elicited priors for that model.

We find that many participants set what they considered to be *weakly informative priors*, a class of prior recommended in the

CHI '20, April 25–30, 2020, Honolulu, HI, USA.
© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376377

Bayesian literature [16, 39]. However, the actual priors these participants set varied widely, suggesting there may not be a common understanding of how to implement them in practice. We discuss ways that normative material on weakly informative priors could be improved through explicitly teaching strategies used by experts in prior setting.

We identify a range of *informativeness* and *skepticism* philosophies that affect how participants approach prior-setting, along with three high-level strategies that researchers use to choose priors: *centrality matching*, *interval matching* and *visual probability mass allocation*. We find that different strategies might be used with different visualizations, so if we show a different visualization people might switch to a better strategy. We discuss how explicitly representing prior philosophies and matching strategies could aid in prior elicitation.

BACKGROUND

A number of recent papers describe advantages of adopting Bayesian inference methods, both in the social sciences [2, 8, 24, 35, 36] and in HCI [31, 42]. Since these arguments have been discussed in detail in previous literature, we instead focus on the role of priors in Bayesian inference, choice of prior types, and the effect of different elicitation techniques.

What is the role of priors in Bayesian inference?

Bayesian inference is a "reallocation of credibility across possibilities", where the "possibilities" are parameters in a statistical model [34]; it consists of declaring our initial assumptions regarding the parameters as probability distributions (priors) and a likelihood function, and using the observed data to update these probability distributions (posteriors).

When large samples of data have been collected, and when the effects that are being estimated are large, the impact of priors on the resulting inferences may not be high. However, if the underlying effects or the sample size is small, prior distributions can have a critical effect on inferences [19]. Yet the choice of a prior for a Bayesian analysis is not always clear, and guidance in how to do so in a principled way varies.

What are the different kinds of prior distributions?

One categorization of prior types in the literature uses *informativeness* to describe priors (Figure 1):

- *Objective or non-informative priors* affect information in the likelihood as weakly as possible; often these are flat, improper priors or bounded, uniform priors.
- *Weakly informative priors* are proper priors which are set up so that the information they provide are intentionally weaker than whatever actual prior knowledge is available; that is, these priors do not take full advantage of domain specific information [18, 46]. These priors try to regularise inferences which are unlikely based on domain knowledge or experiment design. Hence, if the data is sufficiently informative, the likelihood will dominate in the posterior, but if the data is weak, a weakly informative prior will influence the posterior.
- *Informative priors* represent all available relevant information about the problem known before seeing the data.

The choice of prior type is often a philosophical one. Some advocate for non-informative priors to minimize the amount of "subjectivity" injected into the analysis [1, 29]. Others argue that statistical analyses are inherently subjective [12, 17], suggesting that the use uniform or diffuse priors—under the illusion of objectivity—is inappropriate. Gelman et al. [19] advocate for priors that can generate data that is consistent with researcher's understanding of the problem and which yield a model with good predictive performance.

The notion of weakly informative priors [10, 15, 16, 44] has been widely adopted by applied researchers. While Gelman et al. provide a set of principles for setting weakly informative priors [14], it is not known how applied researchers actually put such normative guidance into practice.

How do we elicit probability distributions?

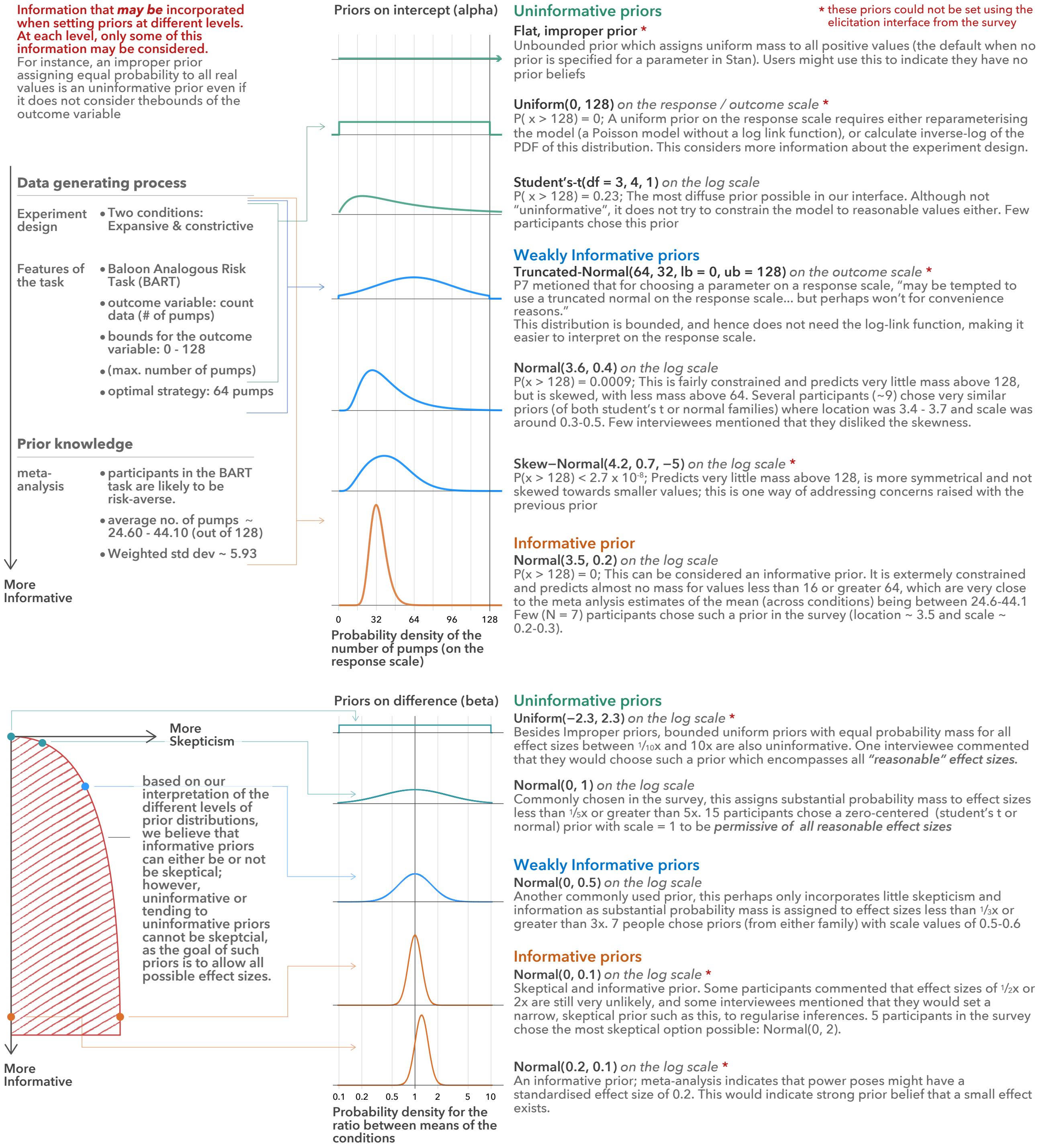
Elicitation of expert[1] knowledge in probabilistic form is a problem with broad applications in fields such as psychology, decision theory, risk assessment, and statistics [41]. Elicitation of expert priors is not easy, and is rarely completely accurate [22, 48]. Good elicitation techniques help researchers represent their prior knowledge as probability distributions which are coherent and, as much as possible, devoid of bias and poor judgement [41]. Several characteristics of the elicitation task can impact the quality of judgements.

Number of variables involved. Eliciting univariate distributions is easier than multivariate distributions, where analysts need to consider joint probabilities for all variables. People exhibit systematic bias when making joint probability assessments [41]. In our study, we present a generalised linear model with two parameters: intercept and mean difference. This is perhaps the most common parameterisation of such models.

Frequency framing. As shown by cognitive psychologists [20, 23], people often find it easier to reason in frequency formats (e.g. 10-in-100 instead of 10%) or in the form of discrete outcomes. In HCI, frequency formats, discrete outcome and probability representations have been applied to improve inferences and decision making from visualizations [11, 21, 25, 26, 30, 32, 33].

Graphical vs textual elicitation. Textual elicitation, which is common, can involve asking an expert for median, and lower and upper quartiles of a distribution [41]. Often more quantiles are elicited, and a parametric probability density function is fitted to the estimates. However, Goldstein et al. [21] show that graphical interfaces—asking users to draw a histogram—can be substantially more accurate than quantile-based methods [38] at eliciting univariate probability distributions. Several studies in HCI have explored different graphical techniques for eliciting prior beliefs from Turkers, either to promote Bayesian reasoning or do Bayesian modeling [32, 33, 49]. Kim et al. [33] found that certain graphical elicitation techniques improved Turkers' Bayesian reasoning when presented with new information. This suggests that asking users to represent their prior beliefs as visualizations can improve the quality of elicitation.

[1] In the elicitation literature, the 'expert' is the person whose knowledge is to be elicited; we use 'user' and 'expert' interchangeably.



Information that *may be* incorporated when setting priors at different levels. At each level, only some of this information may be considered.
For instance, an improper prior assigning equal probability to all real values is an uninformative prior even if it does not consider thebounds of the outcome variable

* these priors could not be set using the elicitation interface from the survey

Uninformative priors

Flat, improper prior *
Unbounded prior which assigns uniform mass to all positive values (the default when no prior is specified for a parameter in Stan). Users might use this to indicate they have no prior beliefs

Uniform(0, 128) *on the response / outcome scale* *
P(x > 128) = 0; A uniform prior on the response scale requires either reparameterising the model (a Poisson model without a log link function), or calculate inverse-log of the PDF of this distribution. This considers more information about the experiment design.

Student's-t(df = 3, 4, 1) *on the log scale*
P(x > 128) = 0.23; The most diffuse prior possible in our interface. Although not "uninformative", it does not try to constrain the model to reasonable values either. Few participants chose this prior

Weakly Informative priors

Truncated-Normal(64, 32, lb = 0, ub = 128) *on the outcome scale* *
P7 metioned that for choosing a parameter on a response scale, "may be tempted to use a truncated normal on the response scale... but perhaps won't for convenience reasons."
This distribution is bounded, and hence does not need the log-link function, making it easier to interpret on the response scale.

Normal(3.6, 0.4) *on the log scale*
P(x > 128) = 0.0009; This is fairly constrained and predicts very little mass above 128, but is skewed, with less mass above 64. Several participants (~9) chose very similar priors (of both student's t or normal families) where location was 3.4 - 3.7 and scale was around 0.3-0.5. Few interviewees mentioned that they disliked the skewness.

Skew−Normal(4.2, 0.7, −5) *on the log scale* *
P(x > 128) < 2.7 x 10^{-8}; Predicts very little mass above 128, is more symmetrical and not skewed towards smaller values; this is one way of addressing concerns raised with the previous prior

Informative prior

Normal(3.5, 0.2) *on the log scale*
P(x > 128) = 0; This can be considered an informative prior. It is extermely constrained and predicts almost no mass for values less than 16 or greater 64, which are very close to the meta anlysis estimates of the mean (across conditions) being between 24.6-44.1 Few (N = 7) participants chose such a prior in the survey (location ~ 3.5 and scale ~ 0.2-0.3).

based on our interpretation of the different levels of prior distributions, we believe that informative priors can either be or not be skeptical; however, uninformative or tending to uninformative priors cannot be skeptcial, as the goal of such priors is to allow all possible effect sizes.

Uninformative priors

Uniform(−2.3, 2.3) *on the log scale* *
Besides Improper priors, bounded uniform priors with equal probability mass for all effect sizes between 1/10x and 10x are also uninformative. One interviewee commented that they would choose such a prior which encompasses all ***"reasonable" effect sizes.***

Normal(0, 1) *on the log scale*
Commonly chosen in the survey, this assigns substantial probability mass to effect sizes less than 1/5x or greater than 5x. 15 participants chose a zero-centered (student's t or normal) prior with scale = 1 to be ***permissive of all reasonable effect sizes***

Weakly Informative priors

Normal(0, 0.5) *on the log scale*
Another commonly used prior, this perhaps only incorporates little skepticism and information as substantial probability mass is assigned to effect sizes less than 1/3x or greater than 3x. 7 people chose priors (from either family) with scale values of 0.5-0.6

Informative priors

Normal(0, 0.1) *on the log scale* *
Skeptical and informative prior. Some participants commented that effect sizes of 1/2x or 2x are still very unlikely, and some interviewees mentioned that they would set a narrow, skeptical prior such as this, to regularise inferences. 5 participants in the survey chose the most skeptical option possible: Normal(0, 2).

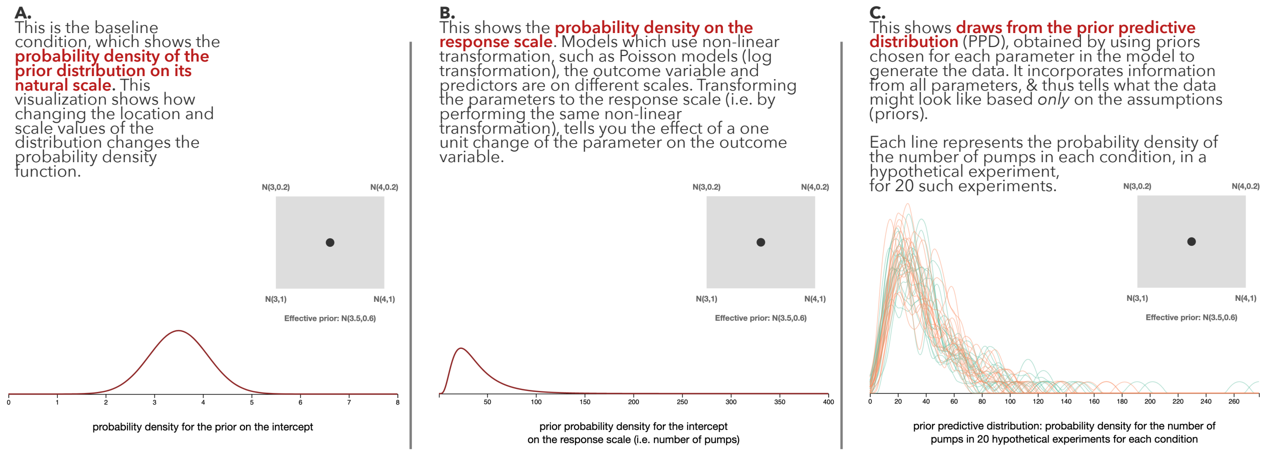
Normal(0.2, 0.1) *on the log scale* *
An informative prior; meta-analysis indicates that power poses might have a standardised effect size of 0.2. This would indicate strong prior belief that a small effect exists.

Figure 1. Information that may be taken into account when different levels of priors are defined. We compare some of the normative levels of priors with what priors participants in our studies have chosen.

METHOD

To perform a descriptive analysis of the prior-setting process we elicit prior distributions for parameters of a statistical model in two separate studies: a survey and a set of interviews. Through the survey we wanted to elicit priors for the described model, understand broad prior setting strategies, and investigate how different visualisations of the prior affect the elicited prior. We then investigated these issues more deeply with specific participants in the interviews.

In both studies, we presented participants with the same model and experiment design. We wanted to probe specific issues common in HCI and applied statistical analyses, such as typical GLM parameterisations and the impact of non-linear link



A.
This is the baseline condition, which shows the **probability density of the prior distribution on its natural scale.** This visualization shows how changing the location and scale values of the distribution changes the probability density function.

B.
This shows the **probability density on the response scale**. Models which use non-linear transformation, such as Poisson models (log transformation), the outcome variable and predictors are on different scales. Transforming the parameters to the response scale (i.e. by performing the same non-linear transformation), tells you the effect of a one unit change of the parameter on the outcome variable.

C.
This shows **draws from the prior predictive distribution** (PPD), obtained by using priors chosen for each parameter in the model to generate the data. It incorporates information from all parameters, & thus tells what the data might look like based *only* on the assumptions (priors).

Each line represents the probability density of the number of pumps in each condition, in a hypothetical experiment,
for 20 such experiments.

N(3,0.2)
N(4,0.2)
N(3,1)
N(4,1)
Effective prior: N(3.5,0.6)
0
1
2
3
4
5
6
7
8
probability density for the prior on the intercept
N(3,0.2)
N(4,0.2)
N(3,1)
N(4,1)
Effective prior: N(3.5,0.6)
50
100
150
200
250
300
350
400
prior probability density for the intercept
on the response scale (i.e. number of pumps)
N(3,0.2)
N(4,0.2)
N(3,1)
N(4,1)
Effective prior: N(3.5,0.6)
0
20
40
60
80
100
120
140
160
180
200
220
240
260
prior predictive distribution: probability density for the number of
pumps in 20 hypothetical experiments for each condition

Figure 2. Different types of visualizations used for prior elicitation

functions. Allowing participants to pick a model from their own work (which we considered) would not guarantee their chosen model would have these properties. Providing a common model was a compromise that allowed us to easily compare priors and prior-setting strategies across participants, with some reduction in ecological validity. To elicit priors, we use three different interactive visualizations of the probability density of the priors. All the materials used in this study, data collected and analysis performed are available online at `https://osf.io/pzu9g/`

Contextual information presented to participants

In both the studies, we presented participants with the design of an experiment conducted by Jansen & Hornbaek [28]:

> *Jansen & Hornbaek performed an experiment to examine the effect of different incidental power poses on risk-taking behavior. They used the Balloon Analogous Risk Task (BART), which is a standard test in Psychology, to measure people's risk-taking behavior in the form of a game. The task was administered through a digital interface. The basic task in BART is to pump up a virtual balloon using on-screen buttons. With each pump, the balloon grows a bit and the player gains a point, which are linked to monetary rewards — the more the players pump up the balloons, the higher their payoff. The maximum size of a balloon is reached after 128 pumps. The risk is introduced through a random, uniformly distributed, point of explosion for each balloon with the average and median explosion point at 64 pumps. The optimal strategy to maximise payoff is to perform 64 pumps. Each participant repeats this 30 times.*

We also provided participants with the results of a meta-analysis of BART studies. Results from meta-analyses can be relevant information that a researcher or analyst might use during prior setting. Participants could use none, some or all of this information while choosing their priors, depending on how informed a prior they might wish to set.

> *A meta-analysis of 22 studies which used the BART task found that the average number of pumps (averaged across conditions) to vary between 24.60 to 44.10 (out of 128 total possible pumps), with a weighted standard deviation of 5.93. This means that based on prior studies, on average, participants in the BART task are most likely to be risk-averse.*

We also described the model to be used for data analysis:

> *The data will be analysed using a Poisson regression model: the outcome variable will be the number of pumps by the participant, and the predictor will be a (categorical) dummy variable indicating which condition the participant is in.*
>
> $$\begin{aligned} pumps_i &\sim Poisson(\lambda_i) \\ log(\lambda_i) &= \alpha + \beta \times condition_i \end{aligned}$$
>
> *where, $condition_i$ has two levels: 0 for the constrictive condition, and 1 for the expansive condition.*

Interactive visualization and prior elicitation

We presented participants with interactive visualizations to aid them in choosing a prior. We visualized the prior in three different ways: *parameter scale density visualization*, *response scale density visualization*, *prior predictive density visualization* (see Figure 2). Although these three visualization types are not exhaustive, *parameter scale density visualization*s and *prior predictive density visualization*s are commonly used for interpreting priors [13]. We have not come across examples of the use of *response scale density visualization*s, but in our example the transformation simply results in a log-normal or log-t distribution with a natural interpretation on the response scale.

We adopted and extended the technique used by Dragicevic et al. [9] to create the interactive visualizations. We explore a different form of elicitation to the techniques than the ones discussed previously, where the user interacts with a widget (Figure 4) to change the location and scale values of a Student's t or a Gaussian distribution—two commonly used prior distributions for parameters in generalised linear models; the corresponding prior is visualized using three different representations (Figure 2). We also provide, in text, the exact location and scale values that the user has set. We chose this elicitation interface because prior probability distributions are commonly parameterised using location (e.g. mean/median) and scale (e.g. standard deviation). The brms statistical modelling package [3] (an R package for specifying Bayesian models), for example, allows the specification of priors with syntax like `normal(0,1)` for a Gaussian prior with mean 0 and standard deviation 1, or `student_t(3, 0, 1)` for a Student's t distribution with 3 degrees of freedom, median 0, and scale 1. Another R package for Bayesian statistical modelling, rstanarm [40], follows a similar syntax.

Survey

Stimuli

The survey was a controlled study with three pages. Visualization type was manipulated between subjects. [2]. The first page (*onboarding page*) introduced participants to the hypothetical experiment design, information from previous studies, and the statistical model; it then introduced participants to the interactive visualizations. Participants had to choose priors for both the parameters in the model: α and β. For each parameter, participants were shown two visualizations (normal and Student's t) and were encouraged to explore the visualizations before proceeding to the second page.

The second page (*elicitation page*) presented participants with two interactive visualizations (one for each parameter) and a *show description* button which expanded a collapsible text container repeating information about the experiment design, previous studies, and the statistical model. On this page, participants set their choice of priors for both the parameters in the model, α and β. For each parameter, they could toggle between the Student's t or normal distributions using a drop-down menu.

The third and final page consisted of eleven fields, of which one was optional. We asked participants to report if they have ever completed a statistical analysis in the past, what statistical software they usually use, their knowledge of this software, their confidence in their choice of priors, and their knowledge of statistics and of Bayesian statistics. We asked participants to describe the strategy used to set the priors, and how they perceived their priors would affect the model. We used these free textual responses and the elicited prior distributions to examine the broad strategies that researchers might be using to specify prior distributions.

Participants

We recruited participants through convenience sampling. Both the authors used Twitter to send out links to the survey (one author maintains software for the visualization of Bayesian results and has a number of followers on Twitter who regularly use Bayesian modelling). We also posted the link to the Transparent Statistics in HCI Slack channel.[3] We did not compensate participants for participation to ensure that participants were not fiscally motivated to participate in our survey and increase the chance that participants were from the target audience: users of Bayesian statistical modelling.

We received a total of 50 responses. All participants reported that they had previously conducted a statistical analysis independently. Figure 3 shows reported levels of experience in different categories. Almost all participants indicated that they were moderate to extremely knowledgeable with statistics and the use of statistical software of their choice; a majority of the participants indicated they were moderate to extremely knowledgeable with Bayesian statistics. This suggests we were able to sample from the desired audience.

Analysis

As preregistered,[4] we wanted to understand the information that was considered while choosing priors and the broad prior setting strategies used, through exploratory and descriptive analysis of the data. Hence, any findings from this analysis should be interpreted qualitatively.

Our analysis method is informed by grounded theory [6]; once the data collection for the survey was completed, the first author used open coding to identify major categories in the data (responses to the free text questions). Whenever the first author was unsure about coding a response, the second author was consulted and the code was discussed until it was resolved; if the response was vague or ambiguous such that neither author could generate a code for it, it was left uncoded. We then clustered the generated codes to identify high-level themes and categories: specifically, *strategies* described by participants and the *consequences* of those strategies (the elicited prior distributions).

Interviews

Procedure and Stimuli

We conducted follow-up interviews with nine survey participants. We followed a semi-structured interview protocol which

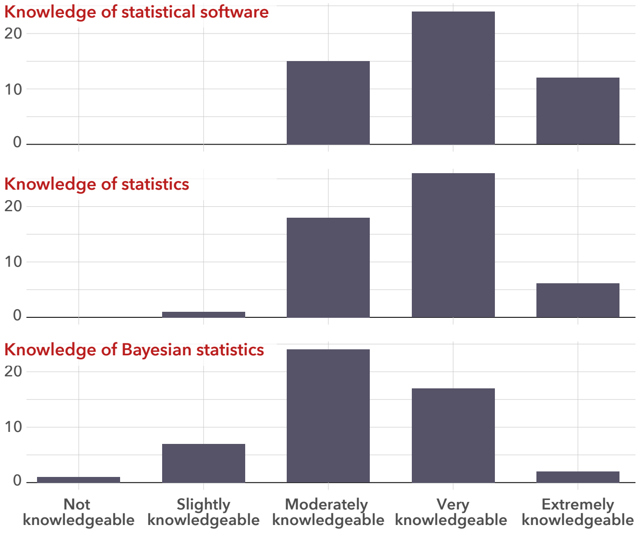


Figure 3. Experience levels of the participants

[2] surveys used in the study can be found in the supplementary materials

[3] https://transparentstatistics.org/

[4] https://aspredicted.org/qw2py.pdf

varied based on participants' responses. First, we provided an overview of the study to the participants. We requested participants share their screen and acquired consent for participation and audio/screen recording. To create a basis for discussion, we created a three-page HTML document for each participant to look at during the interview, and shared an online link to it, requesting participants to open the link.

The first page was the same as the the elicitation page of the survey, showing the same visualizations that participants were shown in the survey. It indicated what prior that participant had set in the survey. This allowed a starting point for the participant to re-acquaint themselves with the visualization and their elicited prior. We prompted participants to think aloud while they were interacting with the visualizations and asked them to walk us through the decisions which led them to their choice of prior (that they chose in the survey).

The second page consisted of the three different interactive visualizations for the mean difference parameter, α. Here, we introduced participants to the two types of visualizations that they had not seen in the survey. Participants were asked to interact with these visualizations and to describe how they would choose a prior distribution for α using each of the two visualizations. Participants were also asked to compare their prior setting process under the different visualizations, and to describe what information they used (or might want to use) when setting their priors.

The third page consisted of the three interactive visualizations for the mean difference parameter, β. As in the second page, participants were introduced to the two visualizations that were unfamiliar and asked to describe how they would choose a prior distribution for β using the two new visualizations.

Analysis

We transcribed the interviews using a professional service. We followed a similar analysis process to that of the survey: the first author used open coding to generate codes from the transcript. The two authors discussed the codes until any discrepancies were resolved. The first author used these codes to find thematic clusters of decision-making strategies, rationales, and information considered. This was done iteratively until we reached a point of inductive thematic saturation [45, 47]: the point when no "new" theoretical insights can be gained from the data [45]. The high-level themes were then discussed by both authors and insights from this process were identified.

RESULTS: SURVEY

Effect of different visualization conditions

Figure 4 presents the priors elicited from the participants in our survey for each condition. Although there appears to be a lot of variation in the priors within each condition, we do not see any substantial differences between conditions. Similarly, we analysed the location and scale values (not shown) chosen by participants and found these to be similar across conditions.

"Weakly informative" means many different things

The normative advice from some prominent Bayesian researchers [18, 19, 46] for choosing priors is to choose "weakly informative priors". While some participants in our survey have explicitly stated that they tried to select weakly informative priors, several others described strategies that we interpreted as closely resembling those for setting a *weakly informative prior*. For example:

> *chose prior for intercept to eliminate density at large number of pumps [...] since I know the balloon will pop before hitting 100 pumps.* [Prior chosen for α: `student_t (3.92,0.21)` in response scale density visualization]

> *Scale of priors was chosen to exclude values greater than 128, with most likely values between 0 and 64.* [Prior chosen for α: `normal(3.52,0.89)` in prior predictive density visualization]

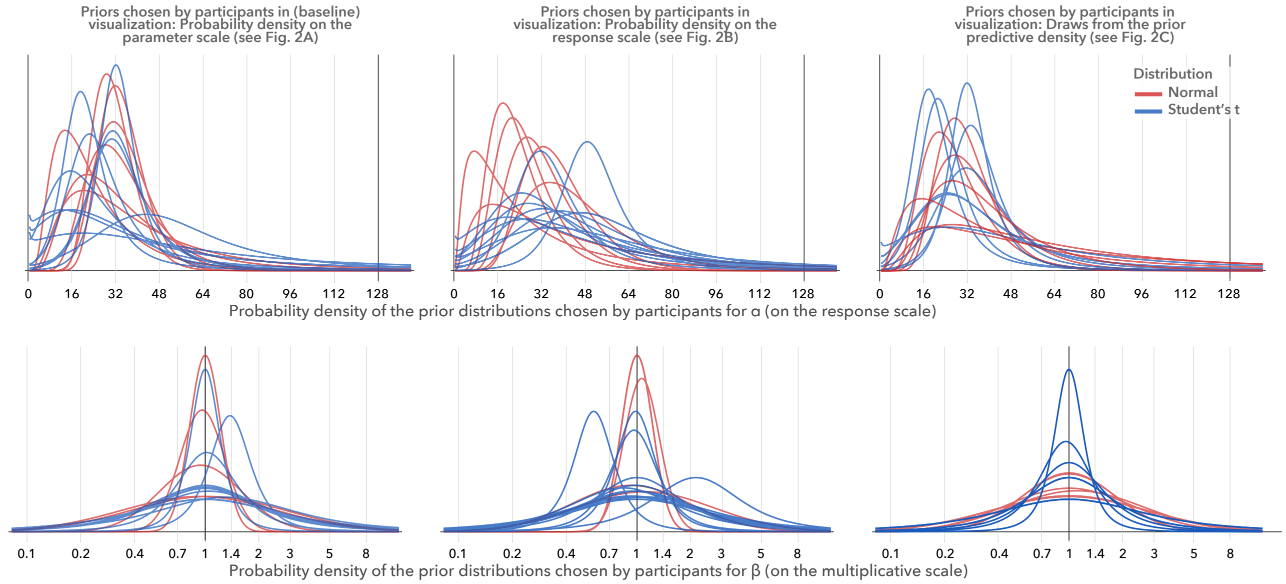


Figure 4. Probability densities of the priors specified by participants in our survey

Based on these descriptions, we coded that 44% (N = 22) of participants tried to select a weakly informative prior for α (the intercept parameter), while 52% (N = 26) tried to select a weakly informative prior for β (mean difference parameter).

One way to interpret Gelman's definition of weakly informative priors is that such priors should minimize the prior probability for theoretically impossible values, which in the scenario presented to participants are values greater than 128. Hence, we calculate the prior predictive probability mass outside the interval [0, 128] by integrating the probability density function for the Poisson process over all possible values of the α parameter (see the supplementary materials for details). We find that the "weakly informative priors" chosen by 68% (15 / 22) of the participants allocated less than 5% prior predictive probability for values greater than 128 (Figures 5B and 5C). However, some of the elicited priors may allocate little or no density at large values, and hence might be considered to be informative rather than weakly informative.

Since most priors elicited for the mean difference parameters were centered around zero (no difference between the two conditions), we compared the values of the scale parameter (Figures 5D & 5E). We find that the "weakly informative" priors chosen by participants encompassed the entire possible spectrum of scale values, indicating many different interpretations of the notion of *weakly informative priors*.

> *setting a prior shape that is wide enough to leave the data drive the posterior sampling, but not too wide as to allow ridiculously absurd values.* [Prior chosen for β: `student_t(-0.04,0.94)` in response scale density visualization]

> *chose distributions that didn't give significant weight to very implausible values.* [Prior chosen for β: `normal (0.06,0.23)` in response scale density visualization]

While one participant applying a "weakly informative" approach chose the smallest value for the scale parameter possible using our elicitation interface (0.2), at least five (out of 26) people chose the largest value for the scale parameter possible using our elicitation interface (1.0) (Figure 5E).

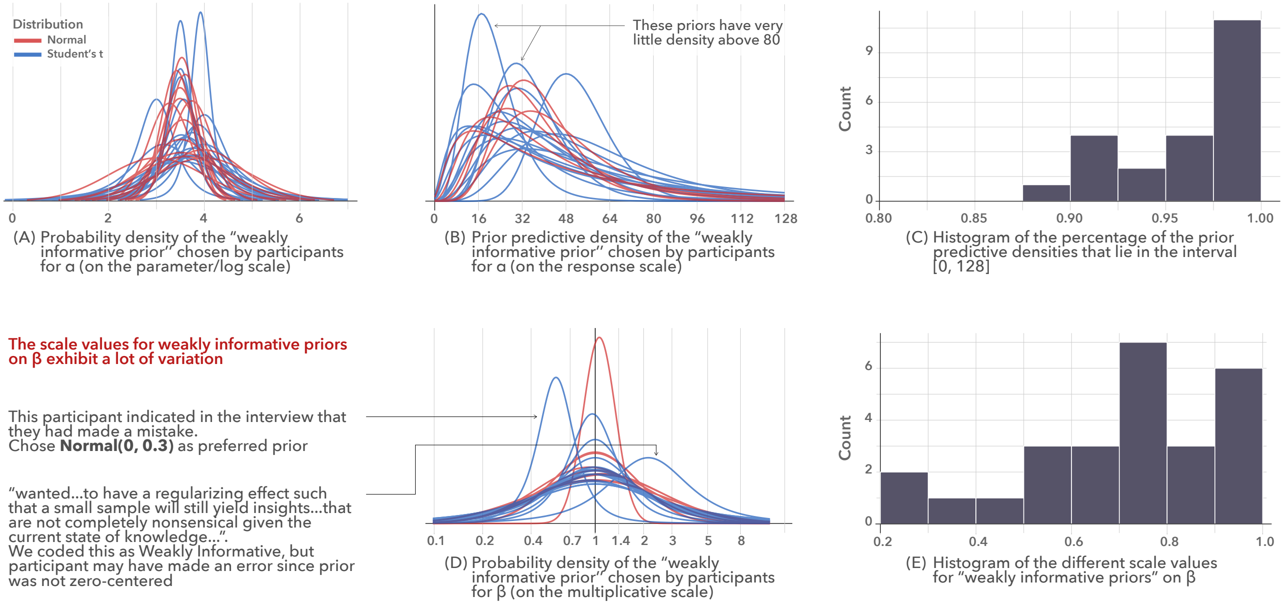
This indicates that although the notion of weakly informative priors is quite popular, there may be quite different interpretations of how to implement such priors in practice. As can be seen from Figure 5D, several of the elicited prior distributions are assigning substantial probability density at effects of 3× or $1/3\times$ in the test condition over the control condition. This may simply be because analysts have different prior expectations of what effect sizes to expect in this context. On the other hand, it may represent misinterpretations of the meaning of effect sizes on a log scale—from our data it is hard to say. What is clear is that the operationalization of "weakly informative prior", particularly for this kind of difference parameter, is not consistent across the analysts who participated in our survey.

RESULTS: INTERVIEWS

Philosophy & experience determine the prior

Researchers can specify prior distributions at different levels of informativeness [14, 46]. Since participants in our study had to specify a proper prior, they could not choose fully "uninformative", unbounded flat distributions. Under those constraints, most participants' elicited prior can be considered either informative, weakly informative, or tending towards uninformative.

Participants' choice of informativeness level was influenced by their statistical ideology and past experience. Their choice of level also affected the extent to which they used the information (such as study design, properties of the BART task, and meta-analysis results) that we presented to them. Broadly speaking, participants fell into the following categories:



(A) Probability density of the "weakly informative prior" chosen by participants for α (on the parameter/log scale)

(B) Prior predictive density of the "weakly informative prior" chosen by participants for α (on the response scale)

(C) Histogram of the percentage of the prior predictive densities that lie in the interval [0, 128]

(D) Probability density of the "weakly informative prior" chosen by participants for β (on the multiplicative scale)

(E) Histogram of the different scale values for "weakly informative priors" on β

Figure 5. Probability densities of the priors we identified as being "Weakly informative"

1. *Those trying to specify an informative prior*. They tried to incorporate information from the meta-analysis into the prior. They generally wanted their intercept prior to have the majority of its probability mass within the interval of 24 to 44—the range provided in the meta analysis—and the expected value for the intercept to be roughly the mid-point of this interval.

2. *Those trying to specify a weakly informative priors*. They usually only considered the experiment design:

> *always worry about cleaving too closely to meta-analyses of past studies, because as we all know, [...] it's better to be slightly wider with the prior than slightly too narrow.*

In general, they wanted to minimize the total probability mass that the prior assigned to values greater than 128, as those values are not theoretically possible.

3. *Those trying to set a prior which tends to uninformative*. They chose the largest possible scale value possible using the interactive visualization. They also expressed that they would rather have set a more diffuse prior than what was allowed.

> *would not be comfortable setting a uniform prior on this, would definitely want a more diffuse prior than what is possible.* — P9 on setting a `normal(0,1)` prior for β.

> *If I don't have any information on the parameters I'm interested in, you just start with the uniform [..] and like just try to set the bounds for something that's reasonable* — P8 on choosing a `student_t(3,0,1)` prior for β.

By choosing the largest possible value for the scale parameter, some participants tended to an uninformative or diffuse prior for β. P8, who usually sets uniform prior, said that they'd rather set a flat prior from [-10, 10] so as to be inclusive of all values they considered plausible.

In the absence of more specific information, past experience influences prior skepticism

Since we did not provide meta-analysis information for the β parameter, we found that participants relied very heavily on their past experience in choosing the prior. Past experience determined their degree of skepticism regarding the presence of an effect, which in turn affected their choice of prior.

> *So my prior would have allowed me up to two times greater. [...] for something like social science that seems like that's pretty big. That still seems pretty permissible* — P1 on choosing a `normal(0,0.2)` prior for β.

> *Just my prior experience with power poses [...] I figured that the effect is not going to be that large.* — P3 on choosing a `student_t(3,0,0.2)` prior for β.

For instance, participants with a background in, psychology, or familiarity with studies in psychology, or with the replication crisis in the social sciences, leaned towards very *skeptical* zero-centered priors: priors with the smallest possible value of scale. In other words, they set their priors such that there would need to be strong evidence for them to conclude that there is a large effect present.

> *one of those situations where you want some amount of expert knowledge, like of what an effect size realistically should be* — P7 on choosing a prior for β.

Some participants wanted to set a prior which did not preclude even large, but probable effect sizes. Participants (P5, P7 and P9) mentioned that they would like to consult an expert regarding what effect sizes are reasonable. However, their field of research, past experience and training appeared to influence what they considered to be large, improbable effect sizes. For instance, P5 and P7 believed that effect sizes of $3\times$ or $\frac{1}{3}\times$ were unlikely, and hence chose a small value for the scale parameter, whereas P8 and P9 wanted to choose more diffuse prior than was possible with the interface:

Student's t as a hedge against mistakes

For both the parameters, participants could choose between a prior from a normal or a Student's t distribution with three degrees of freedom. Participants in general were aware that the Student's t distribution assigned more probability density in the tails compared to the normal distribution. The Student's t distribution is the default prior on the intercept in brms [3], a commonly-used statistical package for Bayesian analysis in R.

> *I want to be close to zero, but I wanted a t distribution because maybe I'm wrong* — P3 on choosing `student_t(3, 0, 0.2)` prior on β.

> *heavy tail on the Student t lets them be a bit wrong on [...] without hurting us too much* — P7 on choosing priors based on information elicited from domain experts.

Participants often chose the Student's t distribution when they expressed a desire to account for the possibility that they were wrong in their choice of prior. On the other hand, participants chose the normal distribution when they did not express a need to accommodate the possibility that they were wrong:e.g., some participants who chose a weakly informative prior for the intercept that allocated most of its mass between 0 and 128 were confident that the intercept will not lie outside that range (impossible given the nature of the task), and so did not feel the need to hedge by using a Student's t distribution.

Different visualizations, different strategies

We found that most participants primarily used a combination of three strategies for determining their choice of priors:

1. *Centrality matching*: trying to to match a central measure (such as mean / median of the prior) to a particular value. For example, if one's prior belief for the expected number of pumps is 50, one might center the prior on this value.

2. *Interval matching*: trying to match the tails of the prior distribution to a particular interval. This was usually observed in elicitation using parameter scale density visualization, as users have a better sense of the 66% or 95% central interval with the familiar bell shape. For example, if one is trying to set a weakly informative prior on the parameter scale, then one might determine an interval, say 0 to 128, where they want approximately 95% of the prior probability to be. They will then calculate the logarithm of those values, and try to match the tails of the visualization so that the desired probability lies within that interval.

3. *Visual probability mass allocation*: trying to visually assess how much mass is allocated above or below specific values. This is an approximate strategy, in contrast to the more precise approach used in interval matching.

The *interval matching* strategy can be implemented in two ways: (1) convert values from the outcome scale to the parameter scale using the logarithm function and then use these to determine the scale parameter of the distribution (based on tail probabilities); (2) calculate the exponent of a value, usually for values around the tails of the distributions, and then determine the tail probabilities assigned to those values.

Visualizations affect which strategy is used

Participants appeared to use different strategies depending on the visualization they used for prior elicitation. Some participants strictly adhered to *centrality matching* and *interval matching* strategies, whereas others used a combination of all three strategies. We found that participants may have a preference towards a certain combination of strategies, which might affect their preferences for particular visualizations.

When priors were elicited through the parameter scale density visualization, participants relied primarily on *centrality matching* and *interval matching*. Here, the probability density was symmetrical and took the form of the familiar bell shape, perhaps making it easier for participants to determine the mean, median, or central 66% or 95% intervals of the distribution, making it easier to implement the *interval matching*. Interpreting distribution on the outcome scale when shown only the parameter scale requires performing exponential transformations, which prevented participants in these conditions from visually inspecting the mass allocation across different values on the outcome scale.

P1 preferred the parameter scale density visualization because the "visual matched the actual prior distribution", and felt that the response scale density visualization added another layer of complexity which they found difficult to use. Only two of the nine participants preferred using the parameter scale density visualization for prior elicitation. We believe that may be because of their strong preference towards the use of the *centrality matching* and *interval matching* strategies.

All other participants said that they disliked setting the prior using the parameter scale density visualization, as they seemed more comfortable thinking of the probability density that the prior assigns to different values on the outcome scale and not on the parameter scale. The parameter scale density visualization required participants to switch between outcome and parameter scales using logarithmic and exponential transformations, which P6 described as "performing mental gymnastics". P7 mentioned that "if you provide logarithms of these numbers, it may be easier to do, but really it's just indirectly getting through that". P7 also mentioned that when eliciting domain knowledge from experts, "you can't ask people what the log mean is, you have to ask what the mean is". Interpreting the probability density of the prior was easier with the response scale density visualization and prior predictive density visualization; the rest of our interviewees explicitly (and strongly) preferred the response scale density visualization. The response scale density visualization allows participants to visually set the prior and inspect probability density assigned to values on the outcome scale; the prior predictive density visualization allows them to see actual draws from the prior predictive distributions. Participants appear to have primarily used a combination of the *visual probability mass allocation* and *centrality matching* strategies in these conditions. Although participants could use multiple strategies when the priors were elicited through the response scale density visualization or the prior predictive density visualization, because the prior distributions were not in the familiar bell shape of the normal or Student's t distributions, using the *interval matching* strategy may have been more difficult.

Transformed coefficients for differences in GLMs can be difficult to reason about

Setting the prior on the β parameter was trickier because it acts multiplicatively; e.g. when $e^{\beta} = 2$, the mean for the test condition would be twice the mean for the control condition. Thus, generalised linear models can be easily misinterpreted. For instance, P9 was "not sure" how to set priors for the model, and hence, was "trying to set it (the prior) to be as diffuse as possible...". Generally, researchers found reasoning in terms of this multiplicative effect difficult when the parameter scale density visualization was used for prior elicitation, but much easier with the response scale density visualization. For example, P5 initially chose their prior as a Normal(0, 1) distribution because, "the tails are not that fat. I don't think the effect is going to be 5 or 3, so the majority of the density is between -2 and 2 which I think would be where the effect might be found". However, they revised the prior when presented with the response scale density visualization, saying "this [...] definitely makes me think a bit more about how the standard deviation of the prior manifests as an effect on the natural scale.".

However, P1 appeared to be more comfortable reasoning about effect sizes on the parameter scale. They were aware that "differences of ±0.5, on the log scale, can be very large" and tend to think more in terms of absolute differences, but found that "thinking of the prior in terms of multiplicity is more difficult". Similarly, P8, who tends to set "uninformative", uniform priors in their own field of work, found the prior elicitation process using the parameter scale density visualization challenging, and was unsure of what prior to choose. Hence, when presented with the response scale density visualization, they found the visualization even more difficult to interpret, partly because they have "never seen the β presented this way".

DISCUSSION

Weakly informative priors are popular but implemented inconsistently

Weakly-informative priors are widely advocated for in the literature and in introductory textbooks [16,39]. Many researchers in our survey claimed to have used *weakly informative priors*, yet the actual priors they chose varied. Some of this variance is likely due to the different backgrounds and experience of participants influencing aspects of their priors (e.g., their level of skepticism); however, it seems likely that at least some of this variance is due to a gap between the high-level notion of weakly informative priors and the particulars of how to implement them in practice. Normative recommendations state that weakly informative priors should be intentionally weaker than whatever actual prior knowledge is available [15,46], but how much weaker and in what way is not always well-defined. The absence of concrete guidelines means the prior depends more on analyst taste and experience, making the task difficult for novices. Normative guidance might be improved by taking advantage of the prior setting strategies we have discussed: novices can be explicitly trained in prior-setting strategies to create weakly-informative priors, like centrality matching, interval matching, or visual probability mass allocation. For example, specific examples of converting domain knowledge about the reasonable range of a parameter into a scale parameter on a normal or Student's t distribution could be incorporated into pedagogical material.

Surfacing informativeness, skepticism, and matching strategies to aid novices in elicitation

Prior-setting strategies could be surfaced in prior elicitation interfaces to help novices. Because people may use different strategies depending on the prior visualization they see, showing certain visualizations might improve prior elicitation by encouraging use of a better strategy. For instance, mass allocation may be a better strategy than interval matching when attempting to constrain priors from unreasonably large values. Since mass allocation is easier to do with a response scale visualization, switching to that visualization might help novices adopt a better strategy.

The deep connection between prior-setting philosophies (informativeness and skepticism), what information to use, and how to use it could also be made explicit in teaching materials. In our experience (corroborated by Phelan et al. [42]), novices to Bayesian inference feel lost when first attempting to set a prior: they do not know where to begin, what information to incorporate, or how to incorporate it. An elicitation interface might first attempt to establish the user's desired high-level approach—such as a weakly informative approach—or indeed help the user discover what approach they wish to take. Such a system could then provide guidance in what information might be useful to set that kind of prior. This information might lead directly to elicitation modes, such as zero-centered priors with interval matching for a straightforward weakly informative prior.

In such an interface, visualizations used for prior elicitation would be designed to explicitly support the strategies necessary to enact the desired prior type. For example, annotations of central tendencies and 50%, 80%, or 95% intervals could help participants with centrality and interval matching strategies. The connection between intervals on a distribution, scale, parameters, and prior knowledge could be encoded directly: e.g., with explicit support for using standard errors from previous studies to set scale parameters on a prior (if an informative prior is desired), or with explicit support for allocating a desired proportion of the distribution to a particular range (perhaps for a weakly informative prior).

Theoretical versus practical flexibility in prior setting

Traditionally, computational efficiency was an important consideration in choosing priors [46], necessitating use of *conjugate priors*, for which posteriors are easy to compute. Yet conjugate priors are restrictive [5]: they prescribe the particular shape of the prior that can be used, which may not reflect the analyst's desired prior. One promise of modern Bayesian methods such as Hamiltonian Monte Carlo (as used by Stan [4]) is that it supports a much wider range of priors without sacrificing computational efficiency. This—*in theory*—means that users are free to select whatever shape of prior best captures their prior knowledge. In practice, we wonder if this promise is regularly fulfilled: while we restricted prior setting in this study to normal and Student's t distributions (thus cannot comment on what other distributions participants might have chosen), few participants seemed to express an interest in looking outside of those prior types. One participant who did so noted that while he might prefer a different prior set on the outcome scale (rather than the log scale), setting such a prior is difficult given the (common) generalized linear model parameterization we used. This presents a trade-off between reparameterizing the model, transforming the desired prior density analytically from the response scale to the log scale, or using a similar (but perhaps less ideal) prior that is easy to specify within the syntax of brms. For novices, the last option is probably the only one available. This suggests that the full technical promise of a wider variety of priors is not yet realized in practice for many users. Elicitation interfaces that aid in simultaneously choosing priors and model parameterizations (or aid in transforming priors between scales) could further unlock the technical promise of modern Bayesian samplers.

Limitations

Because we recruited through our Twitter networks, we sampled largely from users who perform Bayesian analysis with Stan and related R packages such as brms and rstanarm. Weakly informative priors, advocated for by Gelman, may be more common in this community as Gelman is a core member of the Stan development team, as opposed to other experts who may recommend using flat or diffuse priors [34]. We only tested one model type, and though that model comes from a very common model class (generalized linear models) our results may not reflect challenges with other common model types, such as hierarchical models.

CONCLUSION

We identified a variety of prior-setting philosophies and strategies. Our focus on descriptive analysis of prior setting is meant to complement normative advice: this is particularly important given the inconsistent implementation of commonly-recommended weakly informative priors. We hope that our results not only can improve future prior setting recommendations but can be incorporated into interactive prior-setting visualizations for both novices and experts.

REFERENCES

[1] James Berger and others. 2006. The case for objective Bayesian analysis. *Bayesian analysis* 1, 3 (2006), 385–402.

[2] James O Berger and Donald A Berry. 1988. Statistical analysis and the illusion of objectivity. *American Scientist* 76, 2 (1988), 159–165.

[3] Paul-Christian Bürkner and others. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80, 1 (2017), 1–28.

[4] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76, 1 (2017).

[5] James J Chen and Melvin R Novick. 1984. Bayesian analysis for binomial models with generalized beta prior distributions. *Journal of Educational Statistics* 9, 2 (1984), 163–175.

[6] John W Creswell and Cheryl N Poth. 2017. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

[7] Geoff Cumming. 2014. The new statistics: Why and how. *Psychological science* 25, 1 (2014), 7–29.

[8] Zoltan Dienes. 2011. Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science* 6, 3 (2011), 274–290.

[9] Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. 2019. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 65.

[10] Michael Evans, Gun Ho Jang, and others. 2011. Weak informativity and the information in one prior relative to another. *Statist. Sci.* 26, 3 (2011), 423–439.

[11] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. 2018. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Conference on Human Factors in Computing Systems - CHI '18*. DOI: http://dx.doi.org/10.1145/3173574.3173718

[12] Bruno de Finetti. 1974. *Theory of probability: a critical introductory treatment*. Technical Report.

[13] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182, 2 (2019), 389–402.

[14] Andrew Gelman. 2019. Prior Choice Recommendations. (2019). https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

[15] Andrew Gelman and others. 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis* 1, 3 (2006), 515–534.

[16] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian data analysis*. Chapman and Hall/CRC.

[17] Andrew Gelman and Christian Hennig. 2017. Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180, 4 (2017), 967–1033.

[18] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, Yu-Sung Su, and others. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2, 4 (2008), 1360–1383.

[19] Andrew Gelman, Daniel Simpson, and Michael Betancourt. 2017. The prior can often only be understood in the context of the likelihood. *Entropy* 19, 10 (2017), 555.

[20] Gerd Gigerenzer and Ulrich Hoffrage. 1995. How to improve Bayesian reasoning without instruction: frequency formats. *Psychological review* 102, 4 (1995), 684.

[21] Daniel G Goldstein and David Rothschild. 2014. Lay understanding of probability distributions. *Judgment & Decision Making* 9, 1 (2014).

[22] Anthony O Hagan. 1988. *Probability: methods and measurement*. Springer Science & Business Media.

[23] Ulrich Hoffrage and Gerd Gigerenzer. 1998. Using natural frequencies to improve diagnostic inferences. *Academic medicine* 73, 5 (1998), 538–540.

[24] George S Howard, Scott E Maxwell, and Kevin J Fleming. 2000. The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological methods* 5, 3 (2000), 315.

[25] Jessica Hullman, Matthew Kay, Yea-Seul Kim, and Samana Shrestha. 2017. Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty. *IEEE Transactions on Visualization and Computer Graphics* (2017).

[26] Jessica Hullman, Paul Resnick, and Eytan Adar. 2015. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one* 10, 11 (2015), e0142444.

[27] John PA Ioannidis. 2005. Why most published research findings are false. *PLoS medicine* 2, 8 (2005), e124.

[28] Yvonne Jansen and Kasper Hornbæk. 2018. How Relevant are Incidental Power Poses for HCI?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 14.

[29] Edwin T Jaynes. 2003. *Probability theory: The logic of science*. Cambridge university press.

[30] Matthew Kay, Tara Kola, Jessica Hullman, and Sean Munson. 2016a. When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*.

[31] Matthew Kay, Gregory Nelson, and Eric Hekler. 2016b. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*.

[32] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. 2017. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1375–1386.

[33] Yea-Seul Kim, Logan A Walls, Peter Krafft, and Jessica Hullman. 2019. A Bayesian Cognition Approach to Improve Data Visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 682.

[34] John Kruschke. 2014. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

[35] John K Kruschke. 2010. What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences* 14, 7 (2010), 293–300.

[36] John K Kruschke. 2013. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General* 142, 2 (2013), 573.

[37] John K Kruschke and Torrin M Liddell. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25, 1 (2018), 178–206.

[38] Hon-Shiang Lau, Amy Hing-Ling Lau, and Chrwan-Jyh Ho. 1998. Improved moment-estimation formulas using more than three subjective fractiles. *Management Science* 44, 3 (1998), 346–351.

[39] Richard McElreath. 2018. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

[40] Chelsea Muth, Zita Oravecz, and Jonah Gabry. 2018. User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quantitative Methods for Psychology* 14, 2 (2018), 99–119.

[41] Anthony O'Hagan, Caitlin E Buck, Alireza Daneshkhah, J Richard Eiser, Paul H Garthwaite, David J Jenkinson, Jeremy E Oakley, and Tim Rakow. 2006. *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons.

[42] Chanda Phelan, Jessica Hullman, Matthew Kay, and Paul Resnick. 2019. Some Prior (s) Experience Necessary: Templates for Getting Started With Bayesian Analysis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 479.

[43] Martyn Plummer and others. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, Vol. 124. Vienna, Austria., 10.

[44] Nicholas G Polson, James G Scott, and others. 2012. On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 7, 4 (2012), 887–902.

[45] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity* 52, 4 (2018), 1893–1907.

[46] Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G Martins, Sigrunn H Sørbye, and others. 2017. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science* 32, 1 (2017), 1–28.

[47] Cathy Urquhart. 2012. *Grounded theory for qualitative research: A practical guide*. Sage.

[48] Robert L Winkler. 1967. The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical association* 62, 319 (1967), 776–800.

[49] Yifan Wu, Larry Xu, Remco Chang, and Eugene Wu. 2017. Towards a bayesian model of data visualization cognition. In *IEEE Visualization Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe)*.